On the architecture of RNA

NMR structural studies on the substrate hairpin of the Neurospora VS ribozyme and the frameshifting SRV-1 pseudoknot

The work described in this thesis was supported by the Council for Chemical Sciences of the Netherlands Organization for Scientific Research (CW/NWO) and the Katholieke Universiteit Nijmegen (KUN).

ISBN 90-9015269-5

On the architecture of RNA

NMR structural studies on the substrate hairpin of the Neurospora VS ribozyme and the frameshifting SRV-1 pseudoknot

een wetenschappelijke proeve op het gebied van de
Natuurwetenschappen, Wiskunde en Informatica

Proefschrift

ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Nijmegen,
volgens besluit van het College van Decanen
in het openbaar te verdedigen op
dinsdag 11 december 2001,
des namiddags om 1:30 uur precies

door

Paulus Josephus Arnoldus Michiels
geboren op 8 juni 1969 te Boxmeer

Promotores:	Prof. Dr. C.W. Hilbers
	Prof. Dr. C.W.A. Pleij (Universiteit Leiden)
Co-promotor:	Dr. H.A. Heus

Manuscript-commissie:	Prof. Dr. A.P.M. Kentgens
	Dr. W.J.G. Melchers

Contents

Abbreviations		**6**
Chapter 1	Introduction	**9**
Chapter 2	Structure of the ribozyme substrate hairpin of *Neurospora* VS RNA: A close look at the cleavage site	**31**
Chapter 3	New developments in structure determination of pseudoknots	**51**
Chapter 4	Heteronuclear ^{1}H-^{13}C experiments for complete assignment of adenine proton and carbon resonances in $^{13}C/^{15}N$ labeled RNA	**75**
Chapter 5	Solution structure of the pseudoknot of SRV-1 RNA, involved in ribosomal frameshifting	**87**
Appendices		**113**
References		**125**
Summary & Samenvatting		**131**
Curriculum vitae		**141**
Nawoord		**142**

Abbreviations

1D, 2D, 3D	one-, two- and three-dimensional
A	adenine
ATP	adenosine triphosphate
BWYV	beet western yellow virus
C	cytidine
DEPC	di-ethylpyrocarbonate
DNA	deoxyribonucleic acid
DQF-COSY	double quantum filtered correlated spectroscopy
G	guanine
HETCOR	heteronuclear correlation
HDV	hepatitis delta virus
HIV	human immuno-deficiency virus
HMBC	hetero multiple bond correlated spectroscopy
HMQC	hetero multiple quantum correlated spectroscopy
HSQC	hetero single quantum correlated spectroscopy
mRNA	messenger RNA
MD	molecular dynamics
MHz	megahertz
Mg^{2+}	magnesium divalent ion
MMTV	mouse mammary tumor virus
Mn^{2+}	manganese divalent ion
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser enhancement
NOESY	NOE spectroscopy
N	any nucleotide, i.e. A, G, C or U
nt	nucleotide(s)
ppm	parts per million
Phe	phenylalanine
R	purine nucleotide, i.e. A or G

RMSD	root mean square deviation
RNA	ribonucleic acid
RRE	rev-responsive element
rRNA	ribosomal RNA
SLIII	stem-loop III of the antigenomic HDV ribozyme
SRV-1	simian retrovirus type-1
TAD	torsion angle dynamics
TOCSY	total correlation spectroscopy
tRNA	transfer RNA
TYMV	turnip yellow mosaic virus
U	uracil
UTP	uridine triphosphate
VS	varkud satellite

Chapter 1

Introduction

The history of RNA structure determination almost goes back as far as that of the celebrated structure determination of DNA. It took, however, an entirely different course. In fact due to the very nature of RNA and its role in transfer of genetic information one cannot even speak of *the* RNA structure determination. This was not clear at all in the early nineteen fifties when Watson and Crick, then at Cambridge (UK), established the DNA double helix (Watson and Crick, 1953). By that time it had become accepted that DNA was the carrier of the genetic information, but how this could be translated into proteins remained shrouded in darkness. That RNA must be involved in the genetic information transfer had dawned on some people, among whom Watson was perhaps the most outspoken (Watson, 1999). When he moved from Cambridge to Pasadena he teamed up with Alexander Rich to start working on the RNA structure. Unfortunately, the first X-ray crystallography on viral and cellular RNA fibers (Rich and Watson, 1954) did not yield the ordered diffraction patterns that had been observed for DNA. A few years later, after it became possible to enzymatically synthesize RNA with random sequences, new attempts led to similar meager results as obtained for the first X-ray diffraction experiments. However, reflection patterns could be distinguished that looked rather similar to those seen earlier for DNA, and that could be correlated to the ordered helical structures. Later on it was realized that the results were compatible with RNA made up of short double helices connected by stretches of single stranded nucleotides. It was not until 1967 that the first model of an RNA-helix was obtained from X-ray diffraction measurements on fibers of fragmented yeast RNA (Fuller *et al.*, 1967).

The view at that time, that knowledge of the three-dimensional structure of RNA molecules was necessary to be able to understand their function, still holds today. Therefore continued attempts were being made to obtain structural information for RNA. Thus, the last decennia have witnessed extensive structural studies of both RNA and DNA molecules by means of NMR spectroscopy, X-ray crystallography and modeling research. The first high-resolution RNA structure, solved in 1973 by X-ray crystallography, was Crick's "adaptor" molecule, better known as transfer RNA or tRNA (Kim *et al.*, 1973, Robertus *et al.*, 1974). This remained the only natural RNA structure elucidated until the second half of the nineteen nineties. In hindsight it became clear that most larger functional RNA

molecules containing certain flexible elements, most times related to their functionality, were very difficult to crystallize or gave bad NMR spectra. New developments in (enzymatic) synthesis created possibilities for preparing milligram quantities of well-defined RNA fragments. This together with the improvement of NMR and X-ray diffraction technology boosted the structure determination of RNA. Due to the poor dispersion of the RNA proton resonances, in particular those of the riboses, initially only the structures of small RNA hairpins could be determined. However, with new RNA labeling techniques and further technical breakthroughs in the field of NMR spectroscopy much larger molecules became accessible to structural investigations. The largest detailed structure of an RNA molecule solved by NMR spectroscopy (Kolk *et al.*, 1998a) is a 44 nucleotide (nt) fragment forming the 3'-terminus of the turnip yellow mosaic virus (TYMV) RNA genome, which contains a pseudoknot.

After the structure determination of different tRNA molecules the next breakthrough by X-ray crystallography occurred in 1996 when the structure of the 160 nt P4-P6 domain of the Tetrahymena Group I intron was determined by X-ray crystallography. This yielded new insights on long range interactions and revealed novel RNA folding motifs (Cate *et al.*, 1996a). Then, in rapid succession a number of RNA structures were solved. Four years later this trend culminated in the elucidation of the 2.4 Å resolution X-ray structure of the 50S ribosomal subunit of *Haloarcula marismortui*. This study showed that the massive amount of 2833 out of 3045 nucleotides of the large ribosomal subunit together with 27 of its proteins can be characterized in detail (Ban *et al.*, 2000). Considering the fact that this enormous leap forward took a few decades of crystallizing efforts, it is expected that in the near future this will only be exceeded by the elucidation of the high-resolution structure of the full eukaryotic ribosome.

This chapter provides an overview of the forces that direct RNA folding and the different structural features found in naturally occurring RNAs such as messenger RNA, ribosomal RNA, transfer RNA and catalytic RNA, based on what has been learned from the aforementioned structural studies. Also the aptameric RNA molecules (reviewed by Heus, 1997, Herman and Patel, 2000) will be described, since their folding properties largely depend on the same motifs.

The architecture of RNA

RNA molecules are generally single chains composed of nucleotides connected by phosphate groups between the C3'-atom of one residue and the C5'-atom of the next. The building blocks of the four different nucleotides, the ribose moiety and the bases, A, G, C and U, are shown in Figure 1.1. RNA can fold into complicated structures by forming helical domains with canonical and non-canonical base pairs intersected by bulges, hairpin loops and internal loops. Subsequently, the different helical domains may be juxtapositioned such that tertiary interactions are formed predominantly between the different non-canonically base paired regions.

In virtually all structures elucidated by X-ray crystallography or NMR spectroscopy, bases that are not involved in Watson-Crick base pairs most often are hydrogen bonded

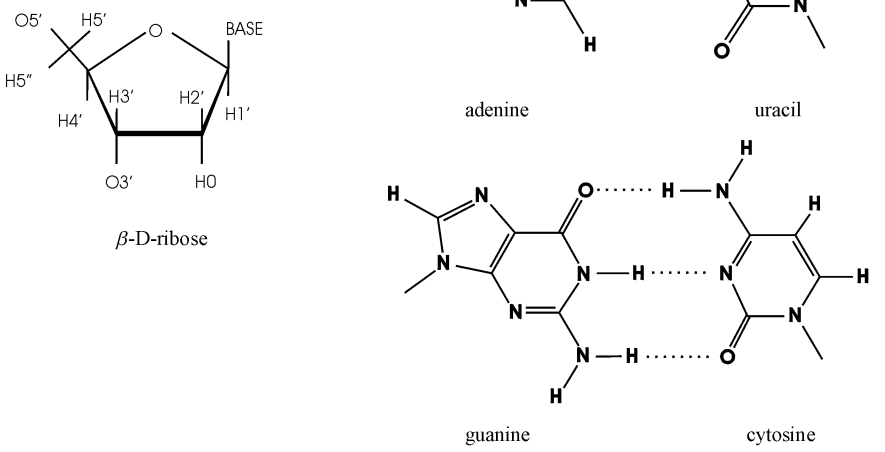

Figure 1.1 *Nomenclature, structure and atom numbering of the β-D-ribose and the four bases (the purines, adenine (A), guanine (G), and the pyrimidines, cytosine (C) and uracil (U) according to IUPAC/IUB guidelines. Dotted lines represent the hydrogen bonds present in the A-U and G-C base pairs.*

and/or stacked. Evidently, there are many more structural features important for RNA structure formation than the A-helical domains. Although RNA consists of only 4 different building blocks (A, G, C and U), its versatility in forming a wide variety of structures is remarkable. Especially, compared with DNA, for which only few structures are known, which essentially only differs from RNA by the absence of the 2'-hydroxyl group. Ultimately, the architecture of a particular RNA molecule, containing alternatively single-, double-, triple- or even quadruple stranded regions, determines its actual functional properties.

Figure 1.2a shows an example of a complicated RNA fold, the secondary structure model of the RNA ribozyme of hepatitis delta virus (HDV). The helical regions stack upon each other resulting in two cylindrical domains formed by the helices I, I.I, IV and II, III, respectively as schematically shown in Figure 1.2b (Ferré-D'Amaré *et al.*, 1998). Two of the central loop nucleotides at the end of helix III, boxed in Figure 1.2a, are base paired to a single stranded region between helical domains I and IV resulting in helix I.I. This is one of the constraints giving rise to the side-by-side orientation of the two cylindrical domains.

In the next paragraphs the principal forces, i.e. the π-π stacking-, the hydrogen bonding interactions and the effect of metal ion binding, that determine the overall fold of RNA will be discussed. Also the recurring motifs found in the present pool of RNA structures, e.g. the U-turn and reversed U-turn, the C-turn, the S-turn, the base- and ribose zipper, the adenosine platform and the A-minor motif, will be reviewed in more detail.

Stacking interactions

From hypochromism measurements (Solie and Schellman, 1968), vapor pressure osmometry (Ts'o and Chan, 1964) and NMR experiments (Schweizer *et al.*, 1968) it was deduced that individual nucleosides stack upon each other at high concentrations in aqueous solution. This base-base interaction appears to be the dominant driving force in RNA folding since for most RNA structures the number of base-base stacking interactions exceeds the number of residues involved in base pairing. For instance, in $tRNA^{Phe}$ (Kim *et al.*, 1973, Kim *et al.*, 1974, Robertus *et al.*, 1974), 42 bases out of 76 are involved in A-type helical domains, though 71 bases are stacked. Purine-purine stacks are the most stable, followed by pyrimidine-purine stacking, while pyrimidine-pyrimidine stacks are the least stable. The order of stability accounts for the observed base stacking in both single- and

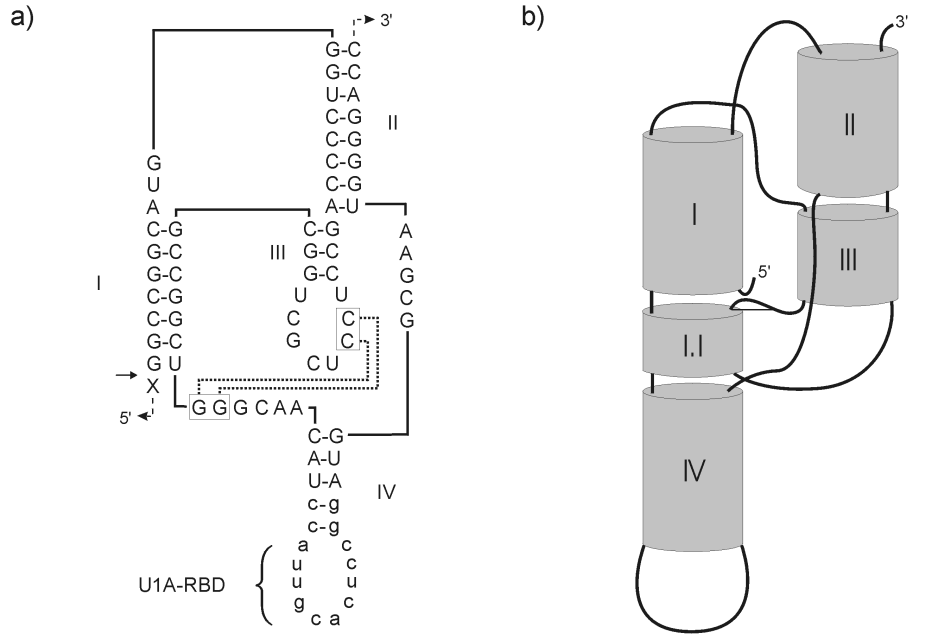

Figure 1.2 *a) Secondary structure model of the genomic HDV ribozyme. The boxed residues are involved in Watson-Crick base pairing comprising helix I.I. Residues in lower case are different from the wild type and comprise the U1A-RNA binding domain (see text). The cleavage site is indicated with an arrow. b) Schematic representation of the different helical domains of the genomic HDV ribozyme structure solved by X-ray crystallography (Ferré-D'Amaré et al., 1998).*

double stranded RNA (Saenger, 1984). Thermodynamic studies showed that single strand stacking interactions of 3'-dangling ends also stabilize RNA structures (Freier *et al.*, 1983, Freier *et al.*, 1985, Burkard *et al.*, 1999). Stacking at the 3'-end occurs in many RNA structures and can even extend up to 3 or 4 nucleotides as was observed in the acceptor stem of tRNA (Kim *et al.*, 1973) and the structure of the TYMV pseudoknot (Kolk *et al.*, 1998a). In our studies of the SRV-1 pseudoknot (Chapter 5) the stabilizing effect of two additional nucleotides at the 3'-end was used to diminish the presence of undesired alternative conformations.

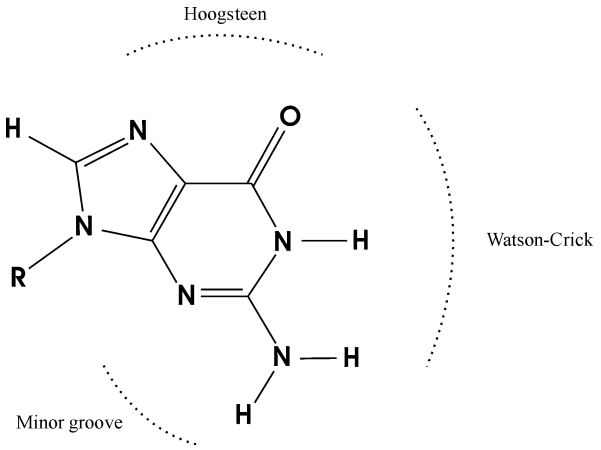

Figure 1.3 *Scheme of the guanine base displaying the three different edges with potential hydrogen bond donors and acceptors involved in base pairing.*

Besides base-base stacking, also base-O4' stacking and base-phosphate stacks have been observed in several structures, for instance in the so-called U-turn and reversed U-turn that will be discussed later.

Base pairing

Within the helical context the Watson-Crick base pairs have the energetically most favorable hydrogen-bonding pattern. A-U base pairs have two hydrogen bonds and G-C base pairs three, thereby being the most stable of the two.

Besides the canonical Watson and Crick base pairs many other hydrogen bond networks can be formed between the donor and acceptor-sites at the Watson-Crick-, Hoogsteen- (major groove) and minor groove edge of the different bases (see Figure 1.3). In RNA the non-canonical G.U wobble and G.A base pairs are most frequently observed. These are occasionally found in the middle of helices but more often occur at the junction of helical domains and in hairpin loops. For instance, the structures of the omnipresent stable RNA UUCG (Cheong *et al.*, 1990) and GCAA tetra-loops (Heus and Pardi, 1991), contain a G.U wobble- and a sheared G.A base pair, respectively, at the junction of the helical stem and the loop region. The variety of mismatched base pairs becomes even larger

if the potential of adenine or cytosine bases to be protonated is included. In this way an additional hydrogen bond donor becomes available. For instance, both the lead dependent ribozyme (Hoogstraten *et al.*, 1998) and the substrate hairpin of the *Neurospora* VS RNA ribozyme, described in Chapter 2, contain an A^+.C base pair.

Mismatched base pairs alter the groove width and backbone conformation compared to the situation in an A-type helical conformation. This may be important for binding of other domains of the RNA molecules (Cate *et al.*, 1996a) or serve as recognition sites for proteins (Szewczak *et al.*, 1993).

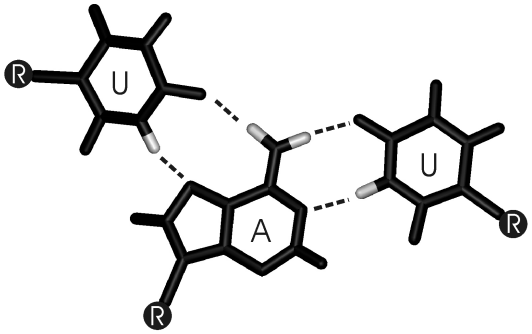

Figure 1.4 *U.A-U triple base pair present in the TAR RNA complexed with the Tat-peptide (Puglisi et al., 1995). Exchangeable hydrogen atoms are colored in grey and the corresponding hydrogen bonds are marked as dotted lines. The ribose units are indicated by the character R*

Base triples

In natural RNA, triple interactions were originally found in tRNA, but also occur in much smaller fragments. A base triple is formed when, dependent on the RNA topology, a third nucleotide forms hydrogen bonds with either the minor or major groove side of a canonical or mismatch base pair. An example of a base triple U.A-U, found in TAR RNA complexed with the Tat-peptide (Puglisi *et al.*, 1995), is shown in Figure 1.4. The third base need not be necessarily donated by a third strand. Major groove triple interactions have also been observed in large internal loops. An example is provided by the large α-sarcin/ricin sensitive loop receptor of 28S rRNA which turned out to be highly structured, with an

internal loop capped by a GAGA tetra-loop (Szewczak *et al.*, 1993). The structure of the internal loop was formed by several non-canonical base pairs and a G.U.A base triple, with a reverse Hoogsteen U.A base pair.

Interactions of a base moiety with the minor groove edge of the RNA base pairs are less frequent. Recently though, a remarkable number of six triple interactions of one of the loops with the minor groove of the (slightly from A-type disturbed) stem were observed in the X-ray crystal structure of the frameshifting pseudoknot of beet western yellow virus (BWYV) (Su *et al.*, 1999).

Base quadruples

Of all base quadruples the G-quartet, formed at high concentrations of guanosine nucleotides in aqueous solutions, is the most well known (Saenger 1984). Earlier results of X-ray fiber diffraction analysis (Tougard *et al.*, 1973) showed the four bases to be hydrogen bonded in a squared manner, hence the name G-quartet, and the base quadruple planes stacked upon each other. This configuration of the guanosine base quadruple was later proposed to exist in the telomeric region of chromosomes (Williamson *et al.*, 1989, Sen and Gilbert, 1988) of which the structure was elucidated both by means of NMR spectroscopy and X-ray crystallography (Smith and Feigon, 1992, Aboul-ela *et al.*, 1994, Laughlan *et al.*, 1994).

Despite the increasing number of solved RNA structures in recent years, only two examples of base quadruple interactions were discovered in RNA. In the structure of the P4-P6 domain of the Tetrahymena Group I intron, the sheared G.A base pair of the GAAA tetra-loop forms a base quadruple with a canonical C-G base pair (Cate *et al.*, 1996a) of the tetra-loop receptor. The other base quadruple, $A.C^{+}.G$-C, discovered in the crystal structure of the BWYV pseudoknot (Su *et al.*, 1999), is differently organized. Here the protonated cytosine is hydrogen bonded to all three other bases. Interestingly, the same base quadruple is also present in the X-ray structure of the HDV ribozyme (Ferré-D'Amaré *et al.*, 1998, Hilbers *et al.*, 1998, see Chapter 3).

Metal binding

Metal ions are important in the overall tertiary fold of RNA structures. Both mono- and divalent cations stabilize RNA by reducing the repulsive forces of the negatively

charged phosphate backbone. This effect is much higher for a divalent than for a monovalent cation. The biologically most important divalent metal ion that binds to RNA is magnesium. In aqueous solutions six water molecules surround the Mg^{2+}-ion. Binding of the magnesium-hexahydrate complex to the RNA can be mediated by hydrogen bond interactions of these water molecules, which is called outer sphere coordination. Alternatively one or more of the water ligands may be replaced by an active group of the RNA molecule, for instance by a phosphate group. In this case the metal ion binds directly to the RNA molecule, which is called inner sphere coordination. In both cases the octahedral coordination is retained and thus six sites can interact with the RNA. The crystal structure of the P4-P6 domain of Tetrahymena Group I intron contains six Mg^{2+}-ion binding sites. One is located in the major groove of a tandem G.U base pair motif near a GAAA tetra-loop. The five other Mg^{2+}-ions are clustered and form a metal ion core in the heart of the ribozyme domain (Cate *et al.*, 1996a, Cate *et al.*, 1997). Single atom mutations engineered to take out one of these magnesium ion-binding sites prohibited folding of the entire P4-P6 domain. Therefore, in analogy to the hydrophobic core in proteins, the magnesium ion core was proposed to act as a nucleation site around which different domains of larger RNA molecules cluster (Cate *et al.*, 1997).

In the crystal structure of the hammerhead ribozyme five Mg^{2+}-metal binding sites are observed (Pley *et al.*, 1994a, Scott *et al.*, 1994). Fluorescence experiments suggested that Mg^{2+} primarily contributes to the stabilization of its tertiary structure and does not cause substantial structural changes (Bassi *et al.*, 1995). This suggests that although Mg^{2+} binds to the hammerhead ribozyme the divalent ions have a non-specific structural role in reducing electrostatic repulsion within the folded RNA. Similar observations were made for RNA pseudoknots where Mg^{2+}-ions can be replaced by Na^{+}-ions, although at much higher concentration, without a change in conformation (van Belkum *et al.*, 1989, Wyatt and Tinoco, 1993).

Metal ions can also be directly involved in chemistry. Mg^{2+}, for instance, is known to act as a cofactor in the cleavage reaction of ribozymes (reviewed by Eckstein and Lilley, 1996). In case of hammerhead cleavage first the hydroxide ion of the $[(OH^-)Mg^{2+}(H_2O)_5]$ complex extracts a proton from the scissile 2'-hydroxyl group. Subsequently, after in-line attack by the ensuing negatively charged 2'-oxygen of the phosphate group, de-

esterification of the backbone results in RNA fragments with 2'-3'- cyclic phosphate and 5'-hydroxyl termini.

Finding specific metal binding sites by means of NMR spectroscopy is difficult, since the binding constants are usually low ($K_a \sim 10^3$-10^4 M^{-1}) and the exchange rates between the bound and unbound states are often unfavorable with respect to the NMR time scale. A qualitative view of metal binding to RNA can be obtained by replacing Mg^{2+} with Mn^{2+}. The paramagnetic Mn^{2+} ion causes line broadening of resonances of nearby groups and therefore indicates the presence of a metal binding site (Bertini and Luchinat, 1986). For RNA this approach has been used by looking at the line broadening of both proton and phosphorus resonances (Allain and Varani, 1995, Hansen *et al.*, 1999).

Cobalthexammine(III), which is similar in size to the Mg^{2+}-hexahydrate complex, has been used several times to investigate the existence of metal binding sites in RNA (Kieft and Tinoco, 1995, Colmenarejo and Tinoco, 1999, Gonzalez and Tinoco, 1999, Rudisser and Tinoco, 2000). In contrast to the water ligands of $Mg(H_2O)_6^{2+}$ the cobalthexammine ligands do not exchange. Consequently it provides for a means to study metal binding mediated via outer sphere coordination. The cobalthexammine-complex that binds to the RNA shows a single line in the NMR spectra for all 18 protons of the hexammine ligands. This suggests that it is freely rotating in the metal binding pockets where it seems to be kinetically trapped. Since the hexammine protons exchange relatively slowly with water protons, intermolecular NOEs to the bound RNA can be observed in the case of favorable binding. Assignment of the proton resonances of the RNA localizes the binding site of the metal complex that can be related to potential magnesium ion binding sites.

Structural motifs

With the increasing number of structures solved by means of NMR and X-ray crystallography recurring motifs, analogous to the α-helices, β-sheets, leucine zippers or zinc-fingers in proteins, are emerging for RNA as well. In retrospect, several of these motifs, such as the base zipper and the U-turn, had already been observed in the X-ray crystal structure of tRNA but at that time were not emphasized as being general structural motifs.

Nearly all motifs discussed in the next paragraphs were observed in the set of structures available prior to the elucidation of the 2.4 Å resolution structure of the 50S ribosomal subunit of *Haloarcula marismortui* (Ban *et al.*, 2000). Many of these are also present in the structures of the 23S and 5S rRNA, which make up the 50S subunit together with several ribosomal proteins. Although so far according to Ban and coworkers (Ban *et al.*, 2000) no novel secondary or tertiary structural motifs have been identified in the 50S subunit, it is very likely that further interpretation of this excessive amount of structural data will in the future reveal novel folding motifs.

U-turn

In $tRNA^{Phe}$, the backbone of the 7-membered anti-codon loop changes direction after the uridine situated 5' to the anti-codon triplet. This motif, which is also present in other RNA fragments, was dubbed the U-turn or π-turn, by analogy with turns in proteins. In general, the U-turn motif involves a three nucleotide sequence 5'-UNR-3' (were U is uracil, N can be any nucleotide and R is a purine) containing a specific hydrogen bond network and base stacking profile as pointed out in Figure 1.5. One hydrogen bond is present between the imino proton of the uridine and the phosphate group at the 3'-side of the purine residue and a second one between the N7 of the purine residue and the 2'-hydroxyl group of the uridine. The sharp π-turn is further stabilized by a base-phosphate stacking interaction. This typical structural feature has also been found in the central core of the hammerhead ribozyme (Pley *et al.*, 1994a, Scott *et al.*, 1995). Here, a conserved 5'-C**UGA**-3' sequence, which forms part of the catalytic pocket, has the signature of a U-turn.

The specific structural features of the U-turn motif are also present within the **GNR**A tetra-loop, although it does not contain the uridine residue (Hilbers *et al.*, 1994, Jucker and Pardi, 1995). Here the guanosine residue donates a hydrogen bond to the opposite phosphate and an identical base phosphate stacking interaction is found. The consequence of the change in direction of the backbone at the turning phosphate is that the residues 3' to the turn protrude into solution available for RNA-RNA interactions. Indeed GNRA loops are frequently found to interact with helices (Pley *et al.*, 1994b) and internal loops, such as in the tetra-loop-receptor motif of the Group I intron (Cate *et al.*, 1996a). In tRNA the U-turn exposes the anti-codon triplet, which is needed for base pairing with the mRNA.

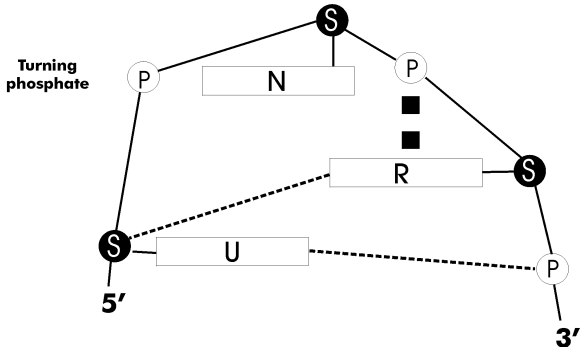

Figure 1.5 *Schematic representation of the 5'-UNR-3' motif, called the U-turn or π-turn. The sugar moieties and the phosphate groups are denoted by the characters S and P, respectively, and the plates represent the bases (uracil (U), any base (N), and purine (R)). The dashed lines correspond with the hydrogen bonds between the 2'-OH of the uridine (U) and the N7 of the purine, and between the imino proton of the uridine and the phosphate group at the 3'-side of the purine. Additionally, the base phosphate stacking interaction between the N and R residue is indicated by the two black squares.*

Reversed U-turn

Like the anti-codon loop of $tRNA^{Phe}$, the central hairpin loop of the anti-genomic HDV ribozyme, stem-loop III (SLIII), also contains seven nucleotides, which are predominantly pyrimidine residues. In this case, however, the phosphate backbone turns in a reverse manner with respect to the U-turn and was therefore called a reversed U- or π-turn (Kolk *et al.*, 1997). In the case of the reversed U-turn a uridine stacks upon the next-neighboring phosphate at the 5'-side, instead of the 3'-side. By comparing the SLIII loop with the central loop in the crystal structure of the genomic HDV ribozyme (Ferré-D'Amaré *et al.*, 1998), it seems likely that the two cytidines 5' to the turning phosphate of the SLIII loop in the anti-genomic HDV ribozyme are involved in a pseudoknot-like interaction with the two guanines nearby the cleavage site. Preformation in RNA folding for RNA-RNA tertiary interaction was clearly demonstrated by Kolk and coworkers for the structure of

both the 3'-hairpin and the full-length 44-nucleotide TYMV pseudoknot (Kolk *et al.*, 1998b). The isolated 3'-hairpin of TYMV turned out to be structured and could practically been superimposed upon the same residues in the cognate pseudoknot structure. Remarkably, also here, stacking of the guanines, that are involved in Watson-Crick base pairing in the pseudoknot structure, occurs at the 5'-side and the direction of the backbone changes at the phosphate of the uridine residue in the top of the loop, similar to the SLIII loop. Like for the 3'-hairpin of the TYMV RNA pseudoknot as well as the U-turn motif, these results also show for the reversed U-turn motif that the base sequence determines the loop structure and the availability of the residues involved in tertiary interactions.

C-turn

Although based upon a single structure - the X-ray structure of the frameshifting pseudoknot of BWYV (Su *et al.*, 1999) - recently the C-turn motif was coined, named after a cytosine preceding the turn. It shares structural similarities with the U-turn and also comprises three nucleotides (5'-CAC-3'). The amino proton of the 3'-cytosine forms a hydrogen bond with the phosphate oxygen 5' to the adenosine. Similar to the U-turn, this is also the second phosphate group although in this case it is counted from the cytosine in the opposite, 5'-direction. One has to realize however, that the C-turn may not be regarded as an independent motif since both cytosines are also involved in Watson-Crick and base quadruple interactions with other residues.

S-turn

While all previous turns reverse the direction of the backbone, S-turns change the course of the backbone locally, while maintaining its overall direction. At the starting point of the turn the sugar moiety has a C2'-endo pucker oriented perpendicular or even anti-parallel with respect to the previous sugar moiety. Directly after the tilted ribose, the backbone direction is reversed. Subsequently the direction of the backbone changes again and continues in the same direction as before giving it an overall S-shape appearance. In the α-sarcin/ricin loop of 28S rRNA (Szewczak *et al.*, 1993), the HIV Rev-responsive element (RRE, Battiste *et al.*, 1996) and in the frameshifting pseudoknot of simian retrovirus type-1 (SRV-1, Chapter 5), the S-turn extends over one or two bases. A seemingly (see Figure 1.6) larger S-turn is observed in the theophylline aptamer (Zimmermann *et al.*, 1997).

However, it actually extends over two residues as well. Starting from the 5'-side, the direction of the backbone is reversed immediately after the last residue involved in an interstrand base-stacking pattern (see next paragraph). After the next two residues, which both have an S-type sugar pucker conformation, the backbone turns back into the original direction.

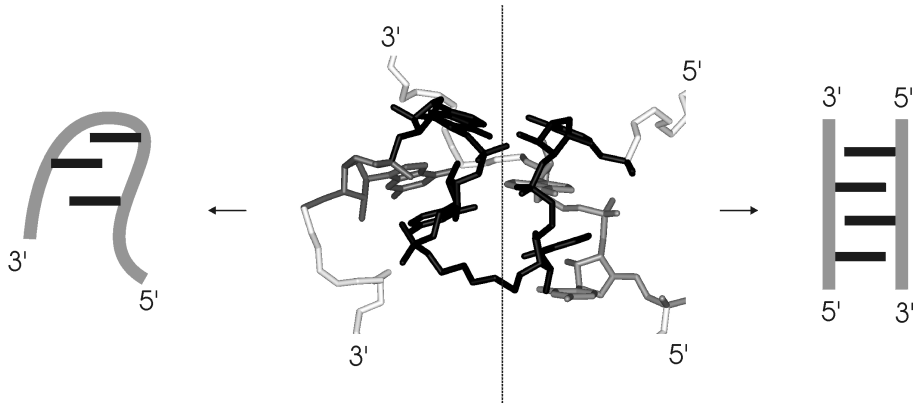

Figure 1.6 *The binding site of the theophylline aptamer (middle) contains an intra- (left) and inter-strand (right) base zipper motif. The intra-strand zipper motif shown on the left is also called a 1-3-2 stack (counted from the 5'-end).*

Base zippers

As mentioned earlier, stacking between bases contributes significantly to the stability of RNA structures and usually occurs in consecutive order within the same strand. However, stacking between different strands has also been observed in several RNA conformations, like in the tandem sheared G.A base pair motif (Santa Lucia *et al.*, 1994, Heus *et al.*, 1997). As for many other motifs, interstrand intercalation of bases was first observed in $tRNA^{Phe}$ between the D- and the T-loop, where two pairs of purines stack in an interleaved manner (Kim *et al.*, 1973, Robertus *et al.*, 1974). This loop-loop interaction involving intercalating bases, together with base pairing interactions, bring about the specific L-shape conformation of tRNA. Base intercalation was also found to stabilize internal loops in the three-dimensional structures of the isolated tetra-loop receptor RNA (Butcher *et al.*, 1997) and loop B of the hairpin ribozyme (Butcher *et al.*, 1999). The term "base zipper", in analogy with the leucine zipper in proteins, was introduced by

Zimmermann and coworkers (1997) upon finding a similar stacking motif in the theophylline aptamer. Besides the intercalation of bases belonging to two different strands, a second mode of base zippering, called the 1-3-2 stack, was found to stabilize the S-turn in the RNA aptamer (see Figure 1.6).

Adenosine platforms

Twenty-three years after the elucidation of the tRNA structure, the high-resolution structure of the P4-P6 domain of the Tetrahymena Group I intron was solved by X-ray crystallography. It exhibited several novel structural features among which the so-called A-A platform (Cate *et al.*, 1996b), important for long-range RNA-RNA interactions, which was one of the most striking novel motifs. In this folding motif the bases of two consecutive adenine residues align within the same plane and form a pseudo base pair on which helical domains and loops of other domains can stack (see Figure 1.7). The crystal structure of the P4-P6 domain contains three such motifs. One is involved in the packing interaction between the two molecules that make up the unit cell. The second widens the minor groove near the P4 and P6 junction, in such a way that triple interactions with another strand can occur. The third A-A platform forms part of the tetra-loop receptor that interacts with the GAAA hairpin loop to align the P5 and P4/6 domains (Figure 1.7). Consecutive stacking is established from the platform throughout all adenines in the hairpin loop. Further stabilization is mediated by hydrogen bonds between the three adenines of the GAAA tetra-loop, including the sheared G.A base pair, and base pairs adjacent to the platform.

The consensus sequence of the platform is not limited to two adenosines, but extends to other base combinations as well. For instance, in the theophylline aptamer an adenosine and a cytosine are aligned similarly (Zimmermann *et al.*, 1997). Therefore a more appropriate name for the A-A platform is the adenosine platform.

Ribose zipper motifs

In a densely packed RNA molecule like the L11-binding domain of 23S rRNA (Conn *et al.*, 1999, Wimberly *et al.*, 1999) or the P4-P6 domain of the Tetrahymena Group I intron (Cate *et al.*, 1996a), the backbones of different domains are in close proximity and riboses can interact via hydrogen bonding involving 2'-hydroxyl groups. This interstrand

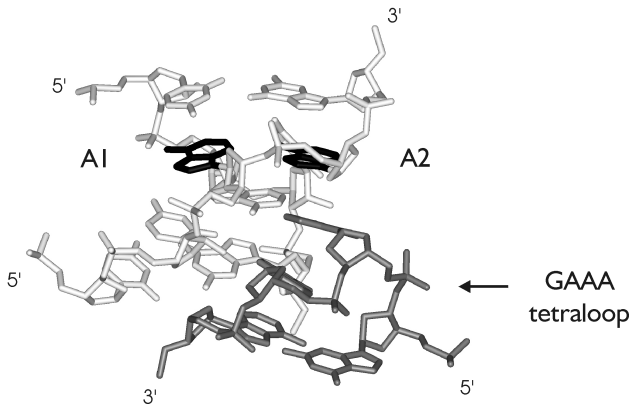

Figure 1.7 *Part of the Group I intron X-ray crystal structure. In A-A platforms two adenine residues, A1 and A2 (colored in black), form a pseudo base pair acting as a platform for RNA-RNA interactions. This figure shows the adenosine platform involved in the tetra-loop receptor - GAAA tetra-loop interaction (colored in dark grey). (Cate et al., 1996a, Cate et al., 1996b).*

interaction, called a ribose zipper, comprises two bifurcated hydrogen bonds from a 2'-hydroxyl group of a residue in one strand to the 2'-hydroxyl group and O2 atom, in the case of a pyrimidine, or N3 atom, in the case of a purine, in the opposite strand (see Figure 1.8). Sofar, the ribose zippers in the hammerhead ribozyme (Pley *et al.*, 1994a), the HDV ribozyme (Ferré-D'Amaré *et al.*, 1998), a double-stranded RNA dodecamer (Schindelin *et al.*, 1995) and in the P4-P6 domain of Tetrahymena Group I intron (Cate *et al.*, 1996a) were all found to exist in pairs. Hermann and Patel (1999) envisaged that even longer ribose zippers could exist in large RNA folds. However, the solution structure of the frameshifting pseudoknot of SRV-1, described in Chapter 5 of this thesis, clearly shows that it can also exist as a single motif.

The A-minor motif

Of all four bases, adenines are by far the most abundant in the unpaired regions of the secondary structure model of 23S rRNA and many of these are conserved (Gutell *et al.*, 2000). This suggests that they can be critical for function or are involved in tertiary interactions stabilizing the overall fold. Examination of the 23S rRNA structure in the 50S

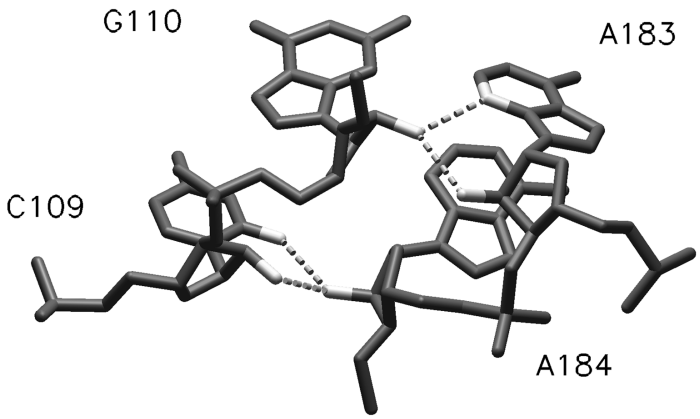

Figure 1.8 *Heavy atom representation of a ribose zipper originating from the X-ray crystal structure of the P4-P6 domain of the Group I intron ribozyme. The dashed lines denote the hydrogen bonds specific for the ribose zipper between the 2'-hydroxyl group and the 2'-hydroxyl group and the N3 and/or O2 in the opposite strand. The numbering of the residues is in accordance to Cate et al. (1996a).*

ribosomal subunit of *H. marismortui* (Ban *et al.*, 2000) revealed that 186 of the 721 adenine residues are involved in RNA minor groove interactions through their N1-C2-N3 edge. This edge of the adenosine allows a tight fit into the minor groove of an RNA helix forming hydrogen bonds with its N1, N3 and hydroxyl group, preferentially with G-C base pairs. In these configurations the phosphate backbone of the adenosine is close to the antiparallel strand of the receptor helix, where they can interact. This typical interaction was recently termed the A-minor motif by Nissen and coworkers (2001). Four different versions were identified, designated type 0-III, depending on the position and the type of hydrogen bonds between the adenine and the minor groove of the helix.

A-minor motifs have actually been observed earlier in other RNA molecules as well. The tertiary interaction of the GAAA tetra-loop with the tetra-loop receptor and the ribose-zipper in the P4-P6 domain of the Tetrahymena ribozyme (Cate *et al.*, 1996a&b),

previously discussed, also include A-minor motifs. Also the loop-helix interaction of the BWYV (Su *et al.*, 1999) and SRV-1 pseudoknots (Chapter 5) are stabilized by these motifs. However, the interaction of the loop-adenine at the junction of the SRV-1 pseudoknot is different from the four types identified by Nissen and collaborators, indicating that there can even be a more extensive repertoire of adenine minor groove interactions in RNA.

Dynamics in RNA, the paradox of structural studies

As already mentioned in the beginning of this chapter, knowledge on the three-dimensional structure of bio-macromolecules could bring us to a better understanding of their working mechanisms. However, there are certain limitations to studying any RNA present in nature. First of all, RNA molecules that adopt multiple conformations cannot be investigated by NMR spectroscopy or X-ray crystallography. Also technical limitations hamper the analysis of RNA. Today, neither NMR spectroscopy nor X-ray crystallography is capable of following the time-dependent course of internal motions in RNA or for that matter in bio-macromolecules. Only the highly ordered rather rigid RNA molecules yield good crystals or behave well as an NMR sample. Therefore, in the current data-set of structures solved by means of NMR spectroscopy or X-ray crystallography, one only finds RNA structures that are well defined with low RMSD values or B-factors, respectively. Rigid molecules, however, can only represent a snap shot from the entire dynamical course of biochemical processes. This is the paradox of studies on structure/function relationships. For instance, despite the elucidation of the X-ray structure of the hammerhead ribozyme (Pley *et al.* 1994a, Scott *et al.,* 1994) the actual cleavage mechanism still remains unclear. The backbone and sugar conformation at the hammerhead's cleavage site are not expected to be rigid, since they have to be rather flexible in order to achieve the proper orientation of the phosphate with respect to the 2'-hydroxyl group for the cleavage reaction to occur. The phosphate group in the crystal structures was not oriented properly towards the 2'-hydroxyl group for the de-esterification reaction of the backbone. It simply means that both X-ray structures reflect an inactive conformation.

Unfortunately, due to the dynamical properties of RNA a whole class of molecules is not amenable for structure determination and as a consequence biological or structural relevant information remains unknown. However, the current set of structures did provide an enormous amount of information on the folding principles and the chemical environment

of structurally or functionally important domains in RNA molecules. In particular, it provides a basis for profound mechanistic studies and gives us a tool to put the results of mutation analysis or chemical modification reactions into the proper perspective of the three-dimensional structure.

Outline of this thesis

At the start of the investigations described in this thesis, all knowledge on RNA architecture was based on the structural information embedded in the tRNA structure and some small hairpin loop- and duplex structures. The discovery that RNA may also possess catalytic properties and the new developments in the field of NMR spectroscopy and X-ray crystallography led to an increased activity in the field, resulting in an enormous expansion of newly solved RNA structures in the last decade. Although it is still very difficult to comprehend or predict the complexity and large variety of structures and associated functions of RNA, considerable progress has been made in understanding the principles of RNA folding. Because of the large number of three-dimensional structures of RNA molecules currently available, increasingly recurring motifs are being recognized in different RNA molecules from different backgrounds and with different functions. However, it is expected that other folding motifs will be found in the future when more complex structures become available. This thesis describes a thorough NMR study of two functionally different RNA molecules, a ribozyme domain and a controlling element of the translational machinery. The studies have lead to a better understanding of the processes they are involved in and have extended our insight in RNA folding.

Chapter 2 involves the first structural studies on domain I of the *Neurospora* VS RNA ribozyme containing the cleavage site. *Neurospora* VS RNA represents a fourth class of ribozymes and differs from the other three (the hammerhead, the hairpin and the HDV ribozyme) in the sense that the substrate strand is not connected to the ribozyme part by canonical base pairing, but is a separate secondary structure element. The substrate hairpin in the *Neurospora* VS RNA ribozyme comprises a symmetrical internal loop within the stem region in which the cleavage site is embedded. The hairpin loop forms a pseudoknot-like interaction with the ribozyme. Although in the wild-type sequence cleavage takes place in *cis*, the substrate can be cleaved in *trans*. Since the smallest active RNA ribozyme fragment of *Neurospora* VS RNA is still too large (~ 120 nt.) for NMR studies, our

structural studies were confined to the internal loop of the substrate. Based on the three-dimensional structure, elucidated by means of NMR spectroscopy, we were able to elaborate on the cleavage mechanism and the possible rearrangements of the cleavage site necessary in de-esterification reaction of the ribozyme.

The remaining Chapters focus on pseudoknot folding. Chapter 3 reviews this typical folding motif, that is found in virtually all classes of RNA playing key roles in a variety of different biological processes (for a review see Deiman and Pleij, 1997). This chapter describes the different aspects and folding topologies of the pseudoknot structure in view of the small number of currently available high-resolution pseudoknot structures. In particular the classical H-type pseudoknot is discussed in depth.

The final two chapters involve the NMR studies on the resonance assignment and structure determination of the H-type pseudoknot involved in frameshifting in the simian retrovirus type-1 (SRV-1) RNA. The assignment of all resonances is the first step in structure determination by means of NMR spectroscopy in order to obtain structural information such as distance restraints from the NOEs, dihedral angle restraints from J-coupling constants or vector angle restraints from dipolar coupling constants. Different from the assignment procedure for proteins using through-bond coupling, standard assignments methods in nucleic acids still heavily rely upon observation of through-space NOEs. It requires foreknowledge about the conformation of the molecule and is therefore less straightforward for residues involved in non-helical regions, such as the two loops of the SRV-1 pseudoknot. Chapter 4 presents two new experiments that can be used for the assignment of all proton and carbon resonances within the adenosine base moiety. The unambiguous assignment of the protons and carbons turned out to be extremely valuable for the structure determination of the RNA pseudoknot involved in ribosomal frameshifting in SRV-1.

Chapter 5 describes the structural studies of the RNA pseudoknot involved in ribosomal frameshifting in SRV-1. As already mentioned, small RNA fragments tend to fold in multiple or alternative conformations. To avoid this, subtle and sophisticated alterations in the sequence had to be made without large deviations from the wild-type and preservation of the functional activity, which is in our case means maintaining the frameshift activity. After several attempts, mutations could be introduced that avoided formation of di- and multi-meric products. Thus, a monomeric variant of the pseudoknot

could be subjected to NMR structural studies. The special structural features found for the loop 2 region explain several results of the mutational analysis, obtained by ten Dam and coworkers. The structural and dynamical properties observed for the junctional region of the frameshifting pseudoknot of SRV-1 are compared with those of other frameshifting pseudoknots and the TYMV pseudoknot.

Chapter 2

Structure of the ribozyme substrate hairpin of *Neurospora* VS RNA: A close look at the cleavage site

Paul J.A. Michiels, Caroline H.J. Schouten, Cornelis W. Hilbers and Hans A. Heus (2000) *RNA*, **12**, 1821-1832.

Abstract

The cleavage site of the *Neurospora* VS RNA ribozyme is located in a separate hairpin domain containing a hexanucleotide internal loop with an A-C mismatch and two adjacent G-A mismatches. The solution structure of the internal loop and helix Ia of the ribozyme substrate hairpin has been determined by nuclear magnetic resonance (NMR) spectroscopy. The two nucleotides in the internal loop, flanking the cleavage site, a guanine and adenine, are involved in two sheared G.A base pairs similar to the magnesium ion-binding site of the hammerhead ribozyme. Adjacent to the tandem G.A base pairs, the adenine and cytidine, which are important for cleavage, form a non-canonical wobble A^{+}-C base pair. The dynamic properties of the internal loop and details of the high resolution structure support the view that the hairpin structure represents a ground-state, which has to undergo a conformational change prior to cleavage. Results of chemical modification and mutagenesis data of the *Neurospora* VS RNA ribozyme can be explained in context with the present three-dimensional structure.

Introduction

Varkud Satellite (VS) RNA is a single-stranded RNA that is found in the mitochondrion of certain natural isolates of *Neurospora.* It contains a self-cleaving RNA motif (see Figure 2.1) that is distinct in secondary structure from other ribozymes studied; the leadzyme, the hammerhead, the hairpin and the hepatitis delta virus (HDV) ribozymes (Beattie *et al*., 1995). However, all these ribozymes generate 2'-3' cyclic phosphate and 5'-hydroxyl termini as cleavage products, suggesting that they are being cleaved via a similar mechanism.

A distinct difference from the aforementioned ribozymes is that the cleavage site in *Neurospora* VS RNA is present in a stem-loop structure, which can be regarded as a separate domain of the ribozyme. In the other cases a helix is formed between the substrate and the ribozyme part by Watson-Crick base pairing at one or both sides of the cleavage site. In VS RNA the substrate hairpin is bound to the much larger ribozyme part through Watson-Crick base pairing between its loop and the loop of stem-loop V in the presence of magnesium (Figure 2.1B, Rastogi *et al*., 1996). The hairpin substrate of *Neurospora* VS

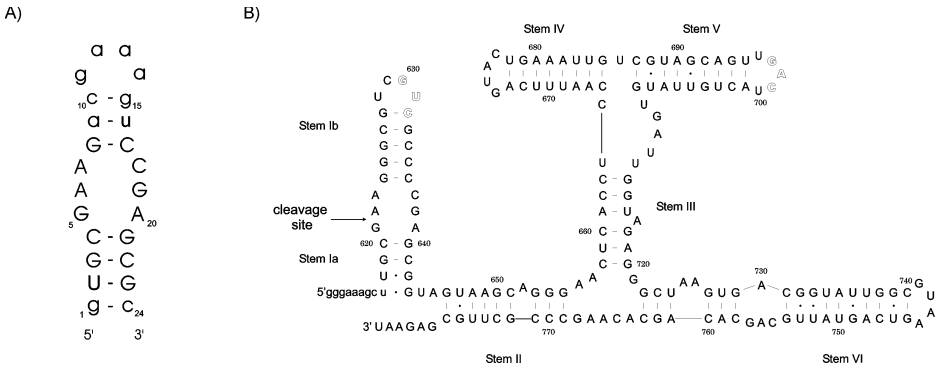

Figure 2.1 *A) Secondary structure of VSm1, the VS substrate hairpin analogue used for the NMR studies. B) Secondary structure of Neurospora VS RNA. Overlapping region of VSm1 and the Neurospora VS RNA is boxed. Non-wild-type residues are given in lower case. The open characters in loop I and loop V are involved in a tertiary interaction.*

RNA can also be cleaved in trans by the remainder of the molecule (Guo and Collins, 1995).

Damage cleavage selection experiments (Beattie and Collins, 1997) using DEPC and hydrazine showed that destabilisation of the internal loop and helix Ia attenuates the cleavage reaction. It was suggested that the structural stability of the internal loop and stem I function as a regulating element tuning the rate of cleavage/ligation of *Neurospora* VS RNA.

Detailed structural information on the different parts of the *Neurospora* VS RNA ribozyme is valuable for understanding how ribozymes function. In this work we have investigated the substrate part of the *Neurospora* VS RNA ribozyme (Figure 2.1A) containing the cleavage site. The results show a hairpin structure with two A-type helical stem regions surrounding an unexpected stable internal loop that consists of three unusual base pairs. A sheared G.A tandem mismatch with an additional A^+-C base pair. The structure of the internal loop nicely explains some of the biochemical data previously conducted, and supports the view that the substrate hairpin represents a ground state that undergoes a conformational change for cleavage to occur.

Results

Figure 2.1A shows the RNA hairpin, termed VSm1, that has been used in our investigation to study the internal loop and helix Ia of the *Neurospora* VS RNA ribozyme. The nucleotides, drawn in upper case, are identical to those in the substrate hairpin of the larger ribozyme (boxed in Figure 2.1B). The stem-loop, involved in tertiary interactions with the stem-loop of helix V, was replaced by a shorter stem and a stable tetra-loop from the GNRA family, as we were primarily interested in the structure of the internal loop. The stable GAAA tetra-loop and the shorter three base pair upper stem simplify resonance assignments and avoid the formation of undesired alternative structures such as dimers. Identical resonance positions for the internal loop were found in both the wild type and the VSm1 hairpin structure (data not shown).

Due to the excellent resonance dispersion of the VSm1 RNA hairpin fragment, nearly complete proton resonance assignment was possible without the aid of isotopically enriched RNA (see chapter Appendices, Table A.I). Heteronuclear NMR experiments on a [^{13}C/^{15}N]-labelled sample, used to study the pH dependence of resonances confirmed the

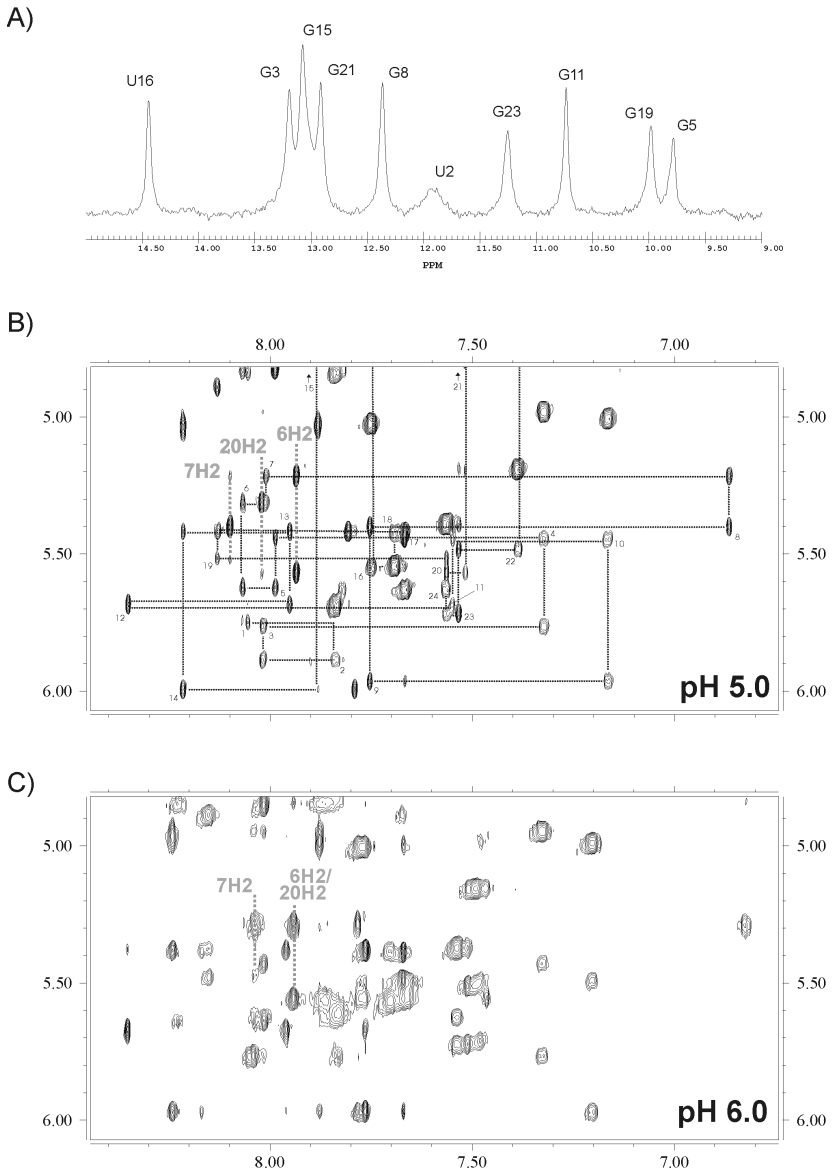

Figure 2.2 *A) 1D ^{1}H NMR spectrum of VSm1 at pH 5.0 showing the imino proton region. B) Detail of a 300 ms NOESY spectrum recorded at 750 MHz at pH 5.0, showing the anomeric to aromatic proton region. Sequential H1'→H8/H6 connectivities are indicated by lines. Intraresidue H1'-H8/H6 NOEs are annotated according to their residue number (cf. Figure 2.1A). C) as (B), recorded at pH 6.0. The H2 protons of the internal loop are marked in grey.*

assignments. Initially, resonance assignment was performed at pH 6.0 (see Figure 2.2C) starting with a canonical anomeric to aromatic proton walk (Wijmenga *et al.*, 1993). Several resonances belonging to residues inside the internal loop though were broadened, possibly by conformational exchange. Lowering the pH to 5.0 (Figure 2.2B) significantly improved the spectral quality through a large decrease in linewidth of these resonances, indicating a higher stability of the internal loop.

At an early stage in this study the main problem in the spectral assignments encountered for the spectra recorded at pH 6.0, was the discrimination between H2 and H8 of A7. In the natural abundance HMQC spectrum (results not shown) no crosspeak was visible at the position of the strongly broadened aromatic H2 and C2 resonances. However, when the pH was lowered to 5.0 the linewidth of this broad resonance decreased by almost a factor of 2 and in the natural abundance HMQC spectrum a crosspeak became visible that was shifted by 7-8 ppm in the carbon dimension. This crosspeak was identified as the H2-C2 to A7.

Subsequently, ^{13}C-^{1}H HMQC spectra at various pH values were recorded to follow the C2 chemical shift of the adenines (see Figure 2.3). Only the C2 resonance position of A7 changed significantly over the measured pH range of 4.5-8. This phenomenon is caused by the protonation of the adenine N1 as has been described earlier for the leadzyme (Legault *et al*, 1994, Hoogstraten *et al.*, 1998) and for loop A of the hairpin ribozyme (Cai and Tinoco, 1996). Like in the leadzyme the protonated and unprotonated species of the VSm1 hairpin are in fast exchange, in contrast to what has been observed for the loop A of the hairpin ribozyme. In the spectra of latter molecule two separate peaks were observed for the protonated and unprotonated adenine. We obtained a pK_a value of 6.4±0.1 which is unusually high, compared to that of free adenine ($pK_a \sim 4$). As will become clear from the structures, the protonated A7 is involved in the formation of an A^+-C base pair in the upper part of the internal loop, similar to what has been found in the leadzyme (Hoogstraten *et al*, 1998) and loop A of the hairpin ribozyme (Cai and Tinoco, 1996). At pH 5.0 A7 is almost fully protonated. Also, no large changes in the imino spectrum were observed by decreasing the pH by one pH unit to 5.0. Because of the decrement in linewidth, the small changes in chemical shifts in the non-exchangeable proton region and the unchanged sequential walk, we concluded that the symmetrical internal loop is stabilised upon protonation, while its

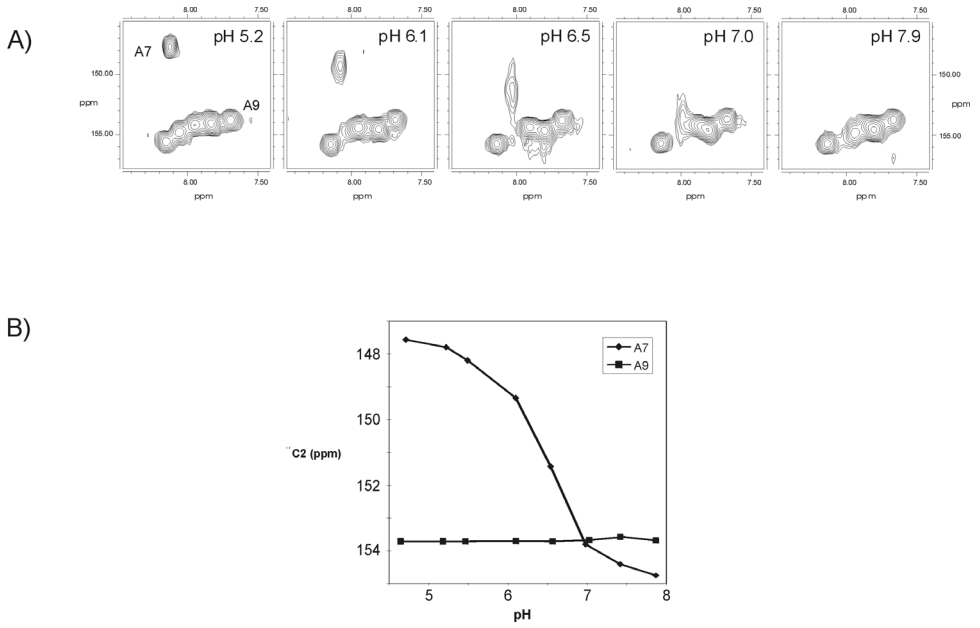

Figure 2.3 *A) H2-C2 region of HMQC spectra of VSm1 recorded at different pH's. Adenine 7 involved in the A^+-C base pair and adenine 9, involved in the A-U Watson-Crick base pair, are marked in the HMQC spectrum at the left hand side. B) Change of C2 chemical shift of adenine 7 compared to that of adenine 9 as a function of pH.*

structure remains unaltered. Therefore further structural studies of the substrate hairpin VSm1 were carried out at pH 5.0.

All imino proton resonances, shown in Figure 2.2A, could be assigned except for G1 due to fraying of the helix terminal end, which also caused broadening of the imino proton resonances of U2 and G23. Broadening was also found for the non-exchangeable imino proton resonances of these residues, again indicating conformational flexibility at the end of the helix. The typical imino proton resonance at 10.75 ppm was assigned to G10 of the GAAA tetra-loop where it is involved in a sheared G.A base pair. Similar to what has been observed by Heus *et al*. (1997) in the UGAA duplex and by Santa Lucia *et al*. (1993) in the CGAG duplex, two far upfield shifted imino protons at 9.98 and 9.77 ppm are present. In

these studies also upfield shifted imino proton resonances that belong to the guanosines involved in a sheared GA mismatches are found.

In the D_2O NOESY spectrum a complete sequential walk could be made from G1 to C24 (see Figure 2.2B). Two unusually high upfield shifted H1' resonances were present, of which the assignments were confirmed by a natural abundance HMQC spectrum. The H1' resonance at 3.62 ppm (see chapter Appendices, Table A.I) belongs to G15 at the 3' side of the GA mismatch in the GAAA tetra-loop and its unusual chemical shift is caused by the ring current effect of adenine 14 that stacks directly above this particular proton. The other upfield shifted H1' resonance (4.15 ppm) was assigned to G21 and its unusual shift is most likely induced by stacking of the neighbouring sheared G.A base pair.

Residues from the GAAA tetra-loop showed the same NOEs and chemical shifts as found by Jucker *et al.* (1996), indicative of identical structures. All NOE connectivities of the two stems and internal loop agree with a regular helical conformation for the stem residues, and a propagation of the helical stacking pattern for all residues in the internal loop. Structurally important NOEs involving the H2s of the three adenines in the internal loop are highlighted in the 2D NOESY recorded in D_2O (Figure 2.2B). Strong NOEs between the A6 H2 to the H1' of A7 and A20, as well as corresponding NOEs between A20 H2 and the H1' of G21 and A6, oriented both adenines. In the 2D NOESY recorded in H_2O important NOEs were observed between the imino proton of G5 and G19 H1' and H8 as well as the G5 amino protons and A20 H8, and from the imino of G19 to G5 H1'(results not shown).

Structure determination

A total number of 298 NOE-derived distance restraints could be collected for the 24-mer hairpin. Inter-residue distance restraints that were collected for the stem regions all supported an A-type helical conformation. For the six internal loop residues 56 inter-residue restraints were obtained, which are schematised in Figure 2.4. No hydrogen bond restraints were imposed on the mismatched base pairs of the internal loop during the structure calculations.

Small H1'-H2' scalar couplings observed in the DQF-COSY and NOESY spectra, were used to infer N-type sugar conformations for the majority of the residues, with the exception of A11 and A12 from the tetra-loop, G5, G19 and the terminal residues G1 and

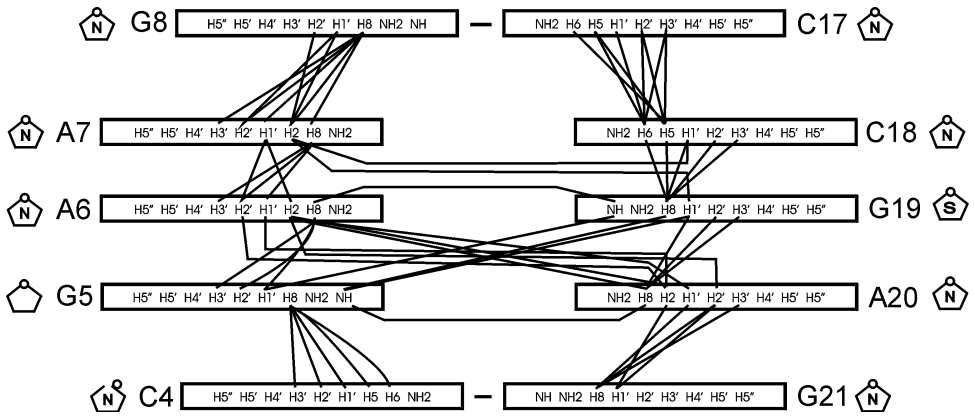

Figure 2.4 *Overview of the inter-residue NOEs, indicated by connecting lines, found for the internal loop region.*

C24. For G19 a large scalar coupling of 8.5 Hz was observed, indicative for a predominantly S-type sugar pucker. The other five residues had scalar couplings of 5-6 Hz, corresponding to sugar pucker rings that rapidly interconvert between N- and S-type conformations. These were therefore left unrestrained during the structure calculations.

More than two third of all structures calculated using the torsion angle dynamic (TAD) approach (Stein *et al*., 1997, Brunger, 1992) converged to a conformation that had no large violations (see Methods). The twenty lowest energy structures were used for further analysis. Figure 2.5A represents the overlay of the twenty lowest energy structures fitted to residues G3-C10/G15-C22; the overlay of the same ensemble fitted to the internal loop is shown in Figure 2.5B. The structural statistics are listed in the chapter Appendices in this thesis, Table A.II.

Discussion

Beattie and co-workers proposed the presence of a number of G.A mismatches at the ends of several helices in the secondary structure model of the large *Neurospora* VS ribozyme, which was based on chemical modification and mutagenesis studies (Beattie *et al*., 1995). G.A mismatches were also proposed for the hexanucleotide internal loop in the

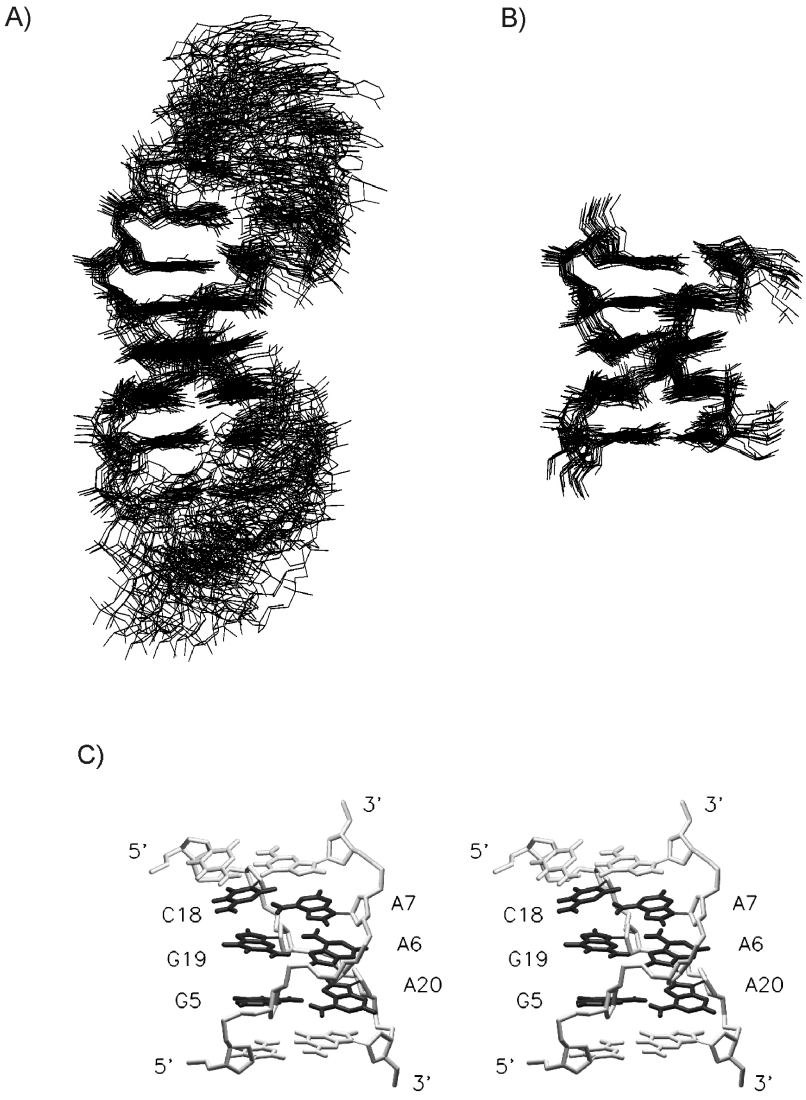

Figure 2.5 *A) Final ensemble of 20 lowest energy structures obtained by superimposing residues G3 through C10 and G15 through C22. B) The same ensemble when the internal loop residues are superimposed; only the internal loop region and the closing base pairs are shown. C) Stereoview of the internal loop residues (colored black) and the closing base pairs.*

substrate hairpin, where the cleavage takes place. As was expected from the secondary structure and the experimental NMR data the structure reveals two small helical stem regions, a typical GAAA tetra-loop and an internal loop with three mismatched base pairs (Figure 2.5C).

Two of the mismatched base pairs form a tandem sheared G.A base pair motif with the scissile phosphate located between residues G5 and A6. The tandem G.A base pair motif shows the typical cross-strand stacking pattern of guanosines and adenines (Figure 2.6), as observed earlier for many other G.A base pair motifs (SantaLucia *et al*., 1993; Pley *et al*., 1994; Scott *et al*., 1995; Heus *et al*., 1997). In each sheared G.A base pair the amino protons of both bases are within hydrogen bond range (< 3Å) of the N3 and N7 of the guanosine and adenine, respectively.

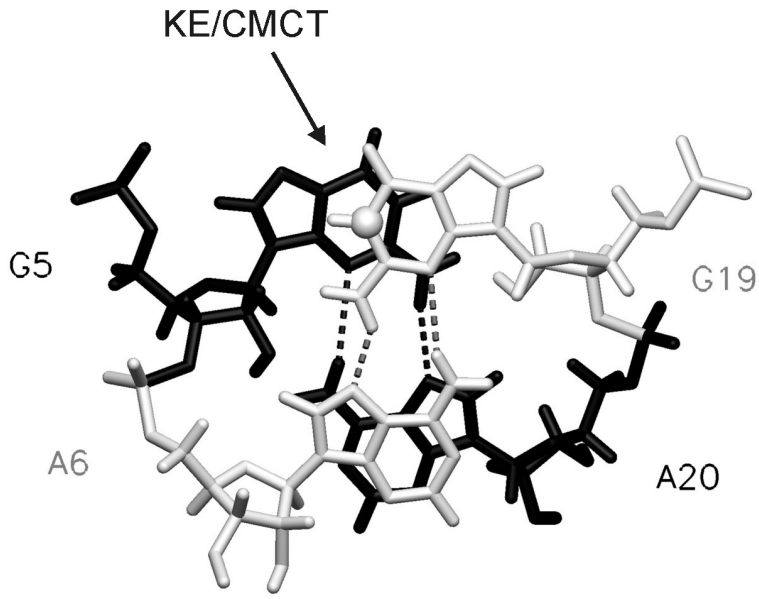

Figure 2.6 *Topview down the helical axis of the tandem sheared GA base pairs in VSm1. Base pair G5-A20 is colored black and base pair A6-G19 is colored grey. The N1 position of G19 that is chemically modified by kethoxal (KE) or CMCT under semi-denaturing conditions and protected under native conditions is indicated as a sphere.*

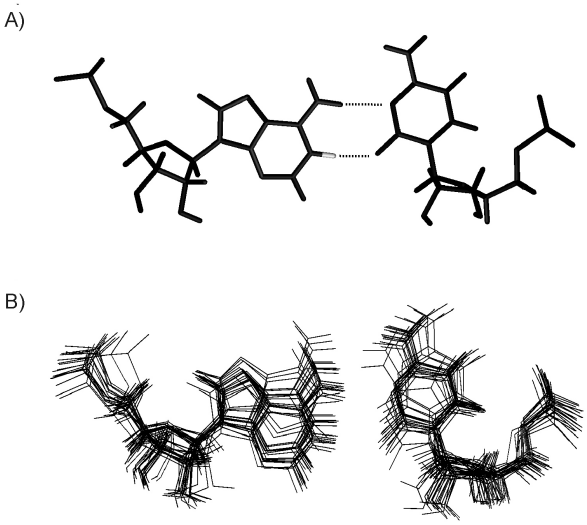

Figure 2.7 *A) Top view down the helical axis of the A^+-C base pair in the lowest energy structure. The unusual imino proton of A7 is colored grey. Hydrogen bonds are shown in dashed lines. B) Superposition of the A^+-C base pairs in the twenty lowest energy structures.*

Furthermore, the NMR structure shows an additional wobble A^+-C base pair at the top of the internal loop, which was not determined by the biochemical studies before. The wobble A^+-C base pair configuration (Figure 2.7) is isosteric with a G-U base pair. In this configuration adenine 7 is protonated at the N1 position, thus stabilizing the A^+-C pair by donating an additional hydrogen bond to the C18 O2. Stabilization of the internal loop by protonation of the N1 agrees with the observed decrease in linewidths of proton resonances upon lowering the pH from 6.0 to pH 5.0. Similar A^+-C base pairing has been observed in several other RNA molecules, such as the leadzyme (Hoogstraten *et al*., 1998), in loop A of the hairpin ribozyme (Cai and Tinoco, 1996) and recently also in an RNA duplex (Pan *et al*., 1998).

As seen in other tandem sheared G.A base pair motifs (Heus *et al*., 1997) in the VSm1 hairpin a large helical twist of 77º between the GA.AG base pairs is compensated by the small helical twists between the sheared G-A base pair and the neighbouring C-G

Watson-Crick base pair on one side and the A^+-C base pair on the other side (17º and 14º, respectively). Also the A^+-C base pair shows the typical small twist (21º) to the base pair on the 3'-side of the adenine, which has been observed earlier for the isosteric G-U base pair (Mizuno and Sundaralingam, 1978; Heus *et al*., 1983). Overall the average helical twist (32º) over the three mismatched base pairs is again nearly equal to an A-form helix (32.7º) changing the helical course only slightly.

Context dependence of tandem sheared G.A base pairs

Based on a structural comparison of three tandem sheared G.A base pair motifs in different contexts Heus *et al*. (1997) suggested that the nucleotide 5' to the guanosine in the tandem sheared G.A base pair motif dictates the sugar pucker conformation, which determines the hydrogen bonding type of the guanosine amino proton in a sheared G.A mismatch. The present structure allows for further comparison and Figure 2.8 shows the hydrogen bond patterns for tandem sheared G.A base pair motifs in four different contexts, of which the structure has been determined so far. When the G.A tandem base pair motif is flanked by U-A base pairs, in the structure termed the *UGAA* duplex, the sugar pucker conformation of the guanosines are found to be of S-type, and the guanosine 2'-hydroxyl forms a hydrogen bond with the opposite guanosine amino group (Heus *et al*., 1997, Figure 2.8A). When the tandem G.A base pairs are flanked by C-G base pairs (in the structure termed the *CGAG* duplex) the guanosines adopt an N-type sugar pucker conformation, and the guanosine 2'-hydroxyl forms a hydrogen bond with the opposite phosphate group (SantaLucia *et al*., 1993, Figure 2.8B). In an asymmetric sheared tandem G.A base pair motif with a C-G base pair on one side and an U-A base pair on the other side, as found in the hammerhead ribozyme structure (Figure 2.8C, Pley *et al*., 1994; Scott *et al*., 1995), the guanosine at the U-A base pair side adopts an S-type sugar, while the guanosine on the C-G base pair side adopts an N-type sugar pucker conformation and the hydrogen bond patterns change accordingly as found in the symmetric G.A base pair motifs.

In the VS substrate hairpin the situation is more complex (Figure 2.8D). Here again the guanosine with an S-type sugar pucker (G19) donates the typical hydrogen bond to the opposite guanosine amino group (G5). Also we found an unusual downfield chemical shift for the phosphorus 5' to A20 (see chapter Appendices in this thesis, Table A.I), indicative for an unusual backbone α- or ζ-angle, as found for the UGAA duplex structure with the S-

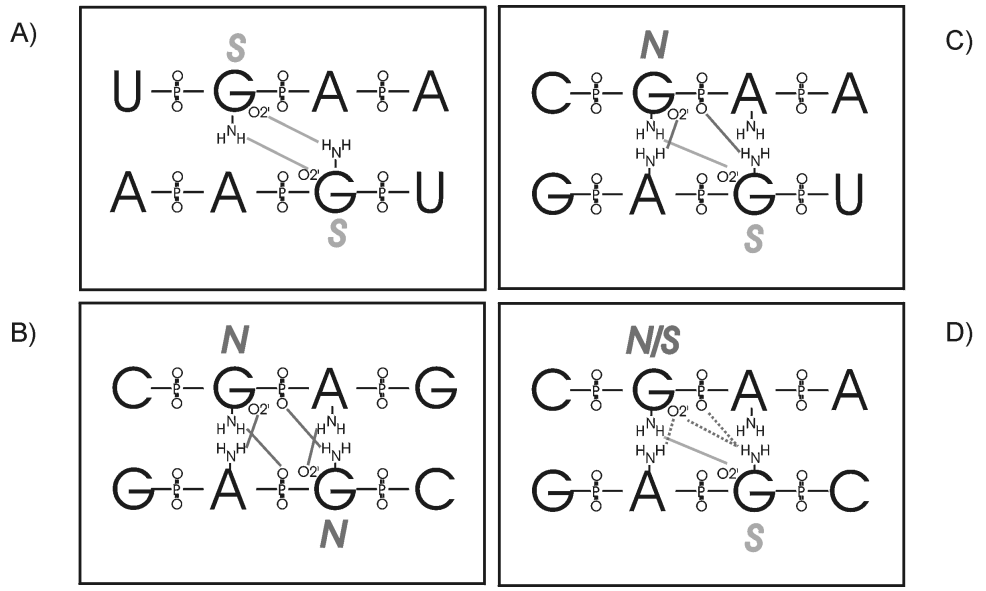

Figure 2.8 *Overview of the hydrogen bonding pattern of four different duplex regions (for details see text) that contain tandem sheared GA base pairs. Sugar puckers of the guanosines are denoted as N- and S-type.*

type guanosine sugar pucker. The presence of an S-type sugar of the guanosine preceded by the cytidine was unexpected, since previously we suggested that the nucleotide at the 5'-side dictates the sugar pucker of the guanosine. Alternatively, the conformational switch from N- to S-type sugar might be related to the lower stability of the C-A^+ mismatch, compared to a closing C-G base pair. On the other side of the GA motif the rapid N/S equilibrium of the G5 ribose obscures the hydrogen bond pattern. Like in the CGAG duplex the ^{31}P chemical shift of A6 was shifted slightly downfield of the region where the phosphates in the helical region resonate. However, in approximately half the number of calculated structures the amino proton of G19 is about equally far away from the O2' as from the phosphate oxygen between G5 and A6, and no direct observation of either hydrogen bond is possible. This does not rule out the presence of the expected hydrogen bonds, but rather hints at a dynamic internal loop, in which these hydrogen bonds are transiently formed. In the calculated structure the backbone between G5 and A6 was very

flexible and for several structures the phosphate was turned towards G19. It is in this orientation that a hydrogen bond can be formed between the phosphate oxygen and the amino proton of G19. The dynamic properties of the internal loop and the two different backbone conformations between residues 5 and 6, where the cleavage site is located might have interesting consequences for the mechanism of cleavage for this ribozyme.

Manganese rescue experiments of phosphorothioate substitutions provided no evidence for a divalent metal binding site in the substrate hairpin (Sood *et al*., 1998). Also, in our magnesium titration studies, apart from some linebroadening, we did not observe specific changes that could arise from magnesium binding (results not shown). At first sight this is highly surprizing, since the NMR structure reveals a G.A base pair motif that bears great similarity with the motif found in the hammerhead ribozyme, where it is associated with a divalent metal ion binding site in the crystal structure (Pley *et al*., 1994, Scott *et al*., 1995,1996). Therefore, analogously to the structure of the hammerhead ribozyme one could expect a divalent metal ion connecting the G19pA20 phosphate-oxygen and the N7 of G20 in the VS substrate hairpin. However, disruption of the closing C-G base pair or removal of helix Ia greatly enhances the rate of cleavage (Rastogi and Collins, 1998), which shows the importance for a less stable internal loop with higher flexibility. Therefore the flexibility, which is already present in the closed VS substrate hairpin might account for the absence of a divalent metal binding site. Absence of this important property upon a very subtle conformational difference illustrates the fine-tuning of RNA architecture and function by RNA folding motifs in different contexts.

Comparison with cleavage rates

Mutation and truncation studies demonstrated the importance of the stability of helix Ia for cleavage (Rastogi and Collins, 1998). Destabilising mutations C4G and G21C increase the cleavage rate of the *Neurospora* VS RNA ribozyme, and removal of helix Ia further increases the cleavage rate. These results suggested that helix Ia is a structural element within the *Neurospora* VS RNA ribozyme that limits the rate of cleavage. Suspiciously, the cleavage rate of the double mutant with a restored G4-C21 base pair is not equal to the wild type level, but remains twice as fast (Rastogi and Collins, 1998). This result suggested that the identity of the closing base pair is important for the stability of helix Ia. It has been shown that tandem G-A base pairs flanked by G-C base pairs, with the

guanosine 5' to the guanosine in the G-A mismatch, adopt a head to head instead of the sheared conformation (Wu and Turner, 1996), and possesses a higher stability than the sheared tandem G.A motif with C-G closing base pairs (Walter *et al*., 1994). Since the high resolution structure reveals a sheared tandem GA motif, changing the closing base pair might force a G-A mismatch into a head to head conformation on one side. However, a combination of both different types of G-A mismatches is geometrically unfavourable, since the C1'-C1' distance across the helix between both types differs by 3.7 Å (9.1 Å for the sheared base pair versus 12.8 Å for the head to head conformation, Gautheret *et al*., 1994). Therefore, fitting in both types of G-A mismatches will probably destabilise the entire internal loop and helix Ia. If this particular mutant substrate hairpin with a closing G-A base pair contains both types of G-A base pair configurations, responsible for the observed lower stability or, alternatively, a tandem sheared or tandem head to head motif accompanied by a lower stability will remain unclear until a high resolution structure of this mutant is available.

Comparison with modification data

From the results of damage selection experiments, the bases of A7, C18 and G19 were identified as being essential for catalytic activity (Beattie and Collins, 1997). Our structure shows that residues A7 and C18 are involved in an A^+-C base pair of which the adenine is partially protonated near neutral pH. From the structure it becomes clear that deletion of the cytidine base by hydrazine, used in the study of Beattie and Collins, will disrupt this base pair. Therefore formation of the A^+-C base pair seems to be essential for cleavage. We also found some higher energy structures with A-C pairs in which instead of the imino proton the amino proton of the adenine is hydrogen bonded to the O2 of the cytidine residue. In this configuration with only one hydrogen bond the N3 position becomes accessible for chemical modification by DMS, which is in line with earlier results of Beattie and co-workers (1995), who observed DMS modification of C18 under semi-denaturing conditions. Under native conditions, in the presence of magnesium, the N3 of C18 is highly protected, which is compatible with the closed A^+-C base pair.

Collins and Olive (1993) found the same cleavage rate of the full VS ribozyme from pH 5.5 through 8.9, and recently Rastogi and Collins (1998) observed a less that twofold rate increase between pH 6 and 8 of a truncated form that lacks stem Ia. In the case of direct

involvement of a hydroxide ion as seen in the hammerhead ribozyme (Dahm *et al*. 1993), one expects a tenfold enhancement of the cleavage rate with each unit increase in pH. Therefore the authors suggested that, most likely, the actual chemical step is not limiting the rate of cleavage, but rather a structural rearrangement prior to cleavage. The pH independence of the full ribozymes does not match the observed pH dependence of A7 in our NMR study. This probably implies that protonation of A7 itself is not important for cleavage, but rather a consequence of formation of the critical A^+-C base pair.

Other modification data compared with our NMR structure also support a conformational change of the substrate hairpin prior to cleavage. G19 can be modified by DEPC under semi-denaturing conditions (Beattie *et al*., 1995). In our structure with the guanosine in the sheared G.A base pair the N7 is still accessible for chemical modification (Figure 2.6). However, under native conditions, i.e. in the presence of magnesium, the N7 is not modified. This suggests that in the active conformation the N7 position is inaccessible and probably involved in hydrogen bonding or protected in some other way. By the same token, the importance of the N7 position of A7, shown by the damage rescue experiments (Beattie and Collins, 1997), points at a similar role of A7 in the active ribozyme. Furthermore recent studies showed an unexpected importance of the phosphate 5' to A7, which is one nucleotide downstream of the scissile phosphate (Sood *et al*., 1998). In our study we see no unusual chemical shift for the A7 phosphorous and in the calculated structures the phosphate backbone between residue A6 and A7 remains A-form. This suggests that this particular phosphate becomes important in the complex with the ribozymal part.

A conformational change prior to cleavage

Thus it appears that the structure of the isolated hairpin represents some kind of a ground state, in which the phosphate backbone does not have the proper orientation for cleavage to occur. Although the exact mechanism of cleavage is unknown, the found reaction products (2',3' cyclic phosphate and 5'-hydroxyl) implicate it involves attack of the 2'-hydroxyl group at the cleavage site (A7) on the adjacent phosphorus atom. The reaction mechanism requires a structural alignment in which the helix Ia is opened such that an in-line conformation of the 2'-oxygen, phosphorus and 5'-oxygen can occur. The biochemical data, combined with the NMR structure support the view that this process is

mediated by a conformational change of the substrate hairpin upon binding to the larger ribozyme part. The three unusual mismatches probably participate in the fine-tuning of the equilibrium between active and inactive conformation, in which the unusual dynamical properties of the backbone around the cleavage site most likely play a key-role. The G-A mismatches in the internal loop of the substrate hairpin might serve as recognition sites for the ribozyme to attack the substrate such that a structural change into the active conformation can occur.

Preformation of secondary structure elements to enable tertiary interactions has been proposed for the TYMV pseudoknot (Kolk *et al*, 1998b). The conformation of the 3'-hairpin loop of the TYMV pseudoknot is very similar to the cognate pseudoknot (Kolk *et al*., 1998a), indicating that the 3'-hairpin is preformed for folding into the pseudoknotted structure. *Neurospora* VS RNA is expected to undergo a much larger conformational change, such as is for instance needed in the case of the GAAA tetra-loop receptor of the Tetrahymena ribozyme (Butcher *et al*., 1997). Whether the structural conformation of the internal loop remains intact and whether they play an active role or act as a nucleation site being recognised by the ribozyme domain of *Neurospora* VS RNA, is still unclear. To address these questions further structural investigations with the substrate hairpin in a larger context with the ribozyme is necessary.

Materials and methods

Sample preparation

An unlabelled and a fully labelled $^{13}C/^{15}N$ 24-mer RNA were prepared enzymatically by in vitro transcription using T7 RNA polymerase (Milligan *et al*., 1987) with synthetic DNA templates. The RNA products were purified using preparative polyacrylamide gel electrophoresis and electroelution. The purified RNA was first deionized using a Biorad Econo-pac column and subsequently lyophilised. The RNA was further washed and concentrated to 500 µl using a Centricon microconcentrator. Final concentrations were 1.8 mM and 1.0 mM for the unlabelled and labelled sample, respectively. Sample buffer conditions were 100 mM NaCl at pH 5.0 unless stated otherwise. Addition of 5 mM $MgCl_2$ did not alter the spectra except for broadening of the resonances and was therefore not further used in data collection.

NMR methods

All spectra were recorded on BRUKER DRX 600 and Varian Inova 500 and 750 MHz spectrometers. Exchangeable protons were assigned from 2D NOESY recorded in 90% H_2O, 10% D_2O at 10° C (Jeener *et al*., 1979) experiments using a jump return sequence (Plateau and Guéron, 1982) or watergate pulsesequence (Sklenár *et al*., 1993c) with a water flip-back pulse (Lippens *et al*., 1995) for water suppression. All other spectra were recorded at 29°C. Non-exchangeable protons were assigned from 2D NOESY spectra recorded in D_2O employing mixing times of 50, 100, 200 and 300 ms, and from DQF-COSY (Shaka and Freeman, 1983) and TOCSY (Griesinger *et al*., 1988) experiments. A natural abundance ^{13}C-^{1}H HMQC experiment (Bax *et al*., 1983) was recorded to discriminate between H2 and H8, and between the H5 of uridine and cytidine residues by the difference in chemical shifts of their attaching ^{13}C. For the determination of the $^{13}C2$ chemical shift dependence with respect of pH, carbon decoupled ^{13}C-^{1}H HMQC experiments were recorded. Phosphorus resonances were assigned by a 2D ^{1}H-^{31}P Hetero-TOCSY-NOESY (Kellog and Schweizer, 1993) and a ^{1}H-^{31}P HETCOR experiment (Sklenár *et al*, 1986).

Structural restraints

Distances involving imino protons were conservatively estimated from NOESY spectra with mixing times of 100 and 200 ms. Other interproton distances were extracted from 2D NOESY spectra with 50, 100 and 200ms mixing times. To account for spin-diffusion effects, distances were calculated predominantly from the 50 and 100ms NOE crosspeak intensities. NOEs involving non-exchangeable protons were converted into distance restraints using the isolated spin pair approximation approach as described by Barsukov and Lian (1993), and the resulting distances were given ±20% error bounds. Scalar couplings between H1' and H2' were determined from a DQF-COSY and a NOESY experiment, and were interpreted as N- or S-type sugar puckers in case of small (<3 Hz) or high (>8 Hz) values, respectively. The sugar puckers of G1, G5, A12, A13 and C24 were left unrestrained because of intermediate scalar couplings of 5-6 Hz. All residues showed H1'/H2'/H3' to H6/H8 NOE intensities that are indicative of an χ-angles in the *anti* range which were therefore restrained to 202±30° for the stem residues and 202±60° for the others.

Both stem regions flanking the internal loop adopt an A-helical conformation, evidenced by chemical shifts, sugar puckers and NOE connectivities. All residues within the stem regions were assigned glycosidic and backbone torsion angles in accordance with canonical A-type RNA with error bounds of ±20°. Hydrogen bonding but no planarity restraints were imposed for all residues involved in Watson-Crick base pairing, evidenced by the imino proton spectra and all observed NOEs and chemical shifts. No restraints were imposed for any of the backbone torsion angles between the sugar of C4 to G8, C10 to G15 and that of C17 to G21, to ensure an unbiased course of the polynucleotide chain. Only the angle ε was set to 225±60°, in order to exclude the stereochemically forbidden *gauche*$^+$ region. The angles α and ζ were restricted to 0±120° if the related ^{31}P chemical shifts resonated within the 1 ppm range typical for A-type RNA.

Structure calculation

A set of 100 structures was calculated using the torsion angle dynamics (TAD) protocol (Stein *et al*., 1997) in X-PLOR (Brünger *et al*., 1992). The TAD cooling step in this protocol was increased to 90 ps so as to obtain a higher convergence rate. Of the obtained pool 68% of the structures had no distance and dihedral restraint violations larger than 0.5 Å and 5°, respectively. Twenty structures of this ensemble, best fitting the experimental data, were selected for presentation, with their structural statistics listed in Table A.II in the Appendices chapter of this thesis. Color figures were generated using MOLMOL (Koradi *et al*., 1996).

Protein Data Bank accession numbers

Coordinates for the final 20 structures have been deposited in the RCSB Protein Data Bank under accession code 1E4P.

Chapter 3

New developments in structure determination of pseudoknots

Cornelis W. Hilbers, Paul J.A. Michiels and Hans A. Heus (1998), *Biopolymers*, **48**, 137-153.

Abstract

Recently, several high-resolution structures of RNA pseudoknots have become available. Here we review the progress in this area. The majority of the structures obtained belong to the classical or H-type pseudoknot family. The most complicated pseudoknot structure elucidated so far is the Hepatitis Delta Virus (HDV) ribozyme, which forms a nested double pseudoknot. In particular, the structure-function relationships of the H-type pseudoknots involved in translational frameshifting have received much attention. All molecules considered show interesting new structural motifs.

Introduction

It is generally assumed that in the folding process of a three-dimensional RNA structure first the secondary elements are formed, after which, through a process of compaction, the final conformation adopts its highly structured shape. In this hierarchic view of RNA folding the first secondary structure element formed is the hairpin or stem-loop. Supporting the view of hierarchic folding is the kinetic data (Crothers *et al.*, 1974, Heus *et al.*, 1983, Hilbers *et al.*, 1985) available for RNA hairpins indicating that they can form with rates as high as 10^5 s^{-1}, while the formation of tertiary interactions are in the millisecond range. The hairpin-stem region can be extended by adding a second helical stem sequentially or simultaneously, which will generate an additional stem separated by an internal loop or a second hairpin. Another possibility is that the bases of the hairpin-loop pair, in a canonical fashion, with complementary bases outside the loop to form a second stem resulting in what has been termed a pseudoknot. The first two additions of stem regions yield classical secondary structures, while in the hierarchy of structures the pseudoknot is considered already a tertiary structure. It can be considered as a stable local (sub)structure which may function as an independent unit or it may be combined from distant parts in a larger functional RNA structure. Particularly in the latter situation, the aforementioned definition of the pseudoknot structure, being formed when bases of the loop pair with a complementary sequence outside the hairpin, may lead to an unclear and ambivalent situation. If in the outside region one involves all kinds of possible structural elements such as hairpin loops, bulges, internal loops or multi-branched loops, and includes long-range interactions, this broad definition is very unsatisfactory in that virtually any tertiary interaction may be viewed as pseudoknot formation. This has caused some debate as to whether pseudoknots defined by long-range interactions – including hairpin loop-loop interactions that are considered tertiary interactions – should be called pseudoknots, or that the term pseudoknot should be restricted to the simpler case where a hairpin loop base pairs with its single stranded tails to form a helical stem adjacent or almost adjacent to the stem of the hairpin.

The “classical” pseudoknot structures or H-type pseudoknots, where H stands for hairpin loop, belong to the latter family. In this family three subclasses with different topologies can be distinguished, one of which is the most abundant (see Figure 3.1). The pseudoknot found in the 3’-end of Turnip Yellow Mosaic Virus (TYMV) RNA (Figure 3.2)

was the first example of this type of molecules (Rietveld *et al.*, 1982). Subsequently, it has been shown that pseudoknots occur in virtually all classes of RNA. They are involved in a variety of biological processes, such as replication initiation, frameshifting, translational control and core formation in larger RNAs (Pleij, 1990, Ten Dam *et al.*, 1992, Pleij, 1994).

Structural characterization of these pseudoknots was mainly conducted by means of chemical modification, enzymatic digestion and sequence comparison. Recently, however,

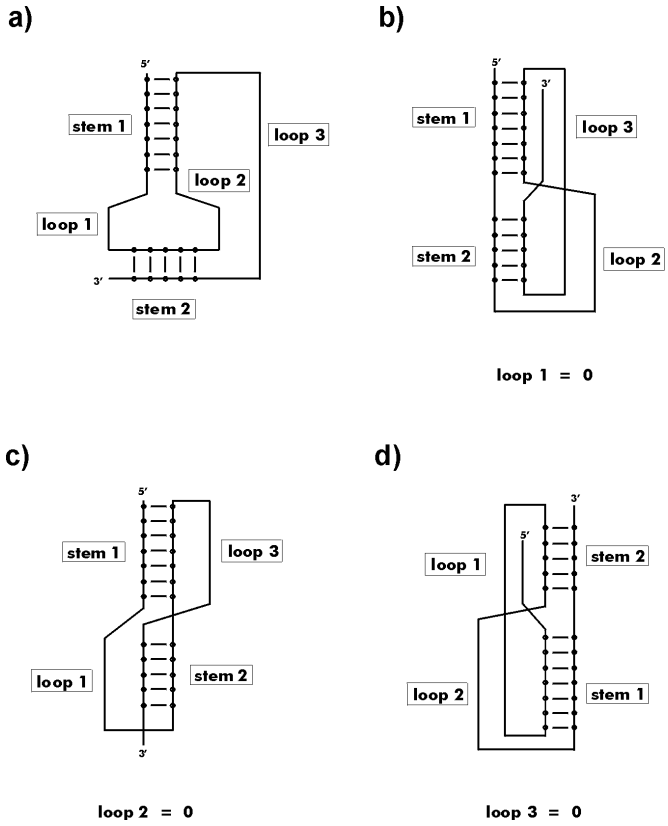

Figure 3.1 *Schematic representation of the classical or H-type pseudoknot. (a) General configuration. (b-d) Configurations in which coaxial stacking of the stems is obtained when one of the loops is reduced to zero. (b) Loop 1 is eliminated. (c) Loop 2 is eliminated (d) Loop 3 is eliminated. The pseudoknot in which loop 2 is eliminated is the one most abundant in natural RNA.*

a number of high-resolution structures obtained via NMR and X-ray diffraction have appeared. All but one of these structures, the hepatitis delta ribozyme, have the topology of the TYMV pseudoknot. In fact to our knowledge no structures with the other classical folding characteristics (Figure 3.1) have as yet been elucidated, although they have been predicted to occur in some molecules, e.g. α-mRNA (Tang and Draper, 1989). These new structures, discussed here, have yielded novel interesting structural motifs, again demonstrating the conformational versatility of RNA. Furthermore, they form the basis for more in-depth studies of structure-function relationships that throw more light on the interesting functions of pseudoknots in cellular processes.

The classical or H-type pseudoknot

In the early 1970s it was discovered that the 3'-end of the RNA genome of many plant viruses can be amino-acylated, suggesting that this part of the RNA folds into a tRNA-like structure (Pinck *et al.*, 1970). Subsequent sequence determination indicated, however, that the RNA chain could not fold into the classical tRNA cloverleaf structure. For one of these plant viruses, the Turnip Yellow Mosaic Virus (TYMV) this is shown in Figure 3.2. The 3'-end of the RNA ends with an $ACCA^{OH}$ sequence, found in a very large number of tRNAs, but this is immediately followed by a hairpin which obstructs the formation of the acceptor stem and thereby the formation of the tRNA cloverleaf structure. It was recognized by Pleij and collaborators (Rietveld *et al.*, 1982) that, through the formation of base pairs between the guanines in the loop of this hairpin and the cytidines upstream in the sequence (see Figure 3.2a), a pseudoknot structure can be generated that then allows the formation of a tRNA-like structure (see Figure 3.2b). Extensive chemical modification and enzymatic digestion experiments established the presence of the pseudoknotted structure in the 3'-end of the TYMV RNA. The stem regions in the pseudoknot were conceived to be stacked coaxially, a necessity to be able to mimic the acceptor stem. Taking into account the right-handedness of the RNA double helix, this meant that loop 1 would cross the major groove of stem 2 and loop 2 the minor groove of stem 1. The possibility of coaxial stacking of the two stems was confirmed later through an NMR study of an H-type pseudoknot formed by a model oligonucleotide (Puglisi *et al.*, 1990).

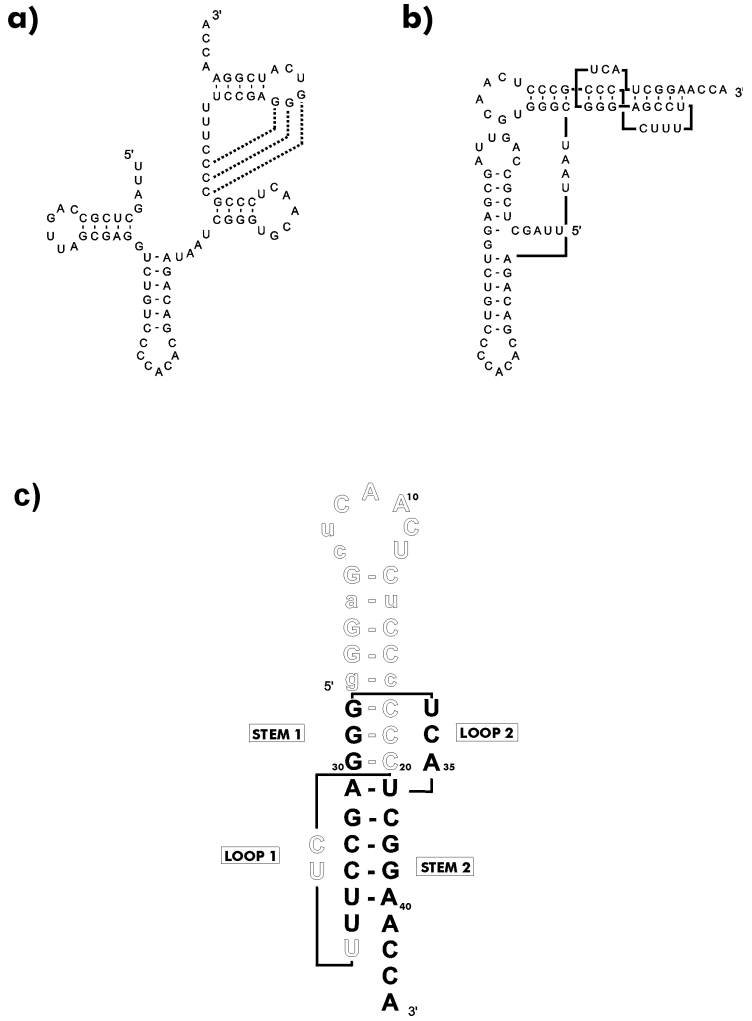

Figure 3.2 *Secondary structure models of the 3'-end of TYMV RNA. a) Secondary structure proposed by Rietveld et al. (1982) and Florentz et al. (1982). The dotted lines represent base pairing between unpaired regions suggested by Pleij and collaborators that leads to the formation of a pseudoknot. b) L-shaped, tRNA-like structure that can be formed upon the introduction of the pseudoknot in the acceptor stem. c) The 44-nucleotide RNA fragment representing the T-arm and the pseudoknotted acceptor-stem of TYMV RNA. This fragment was used to elucidate the pseudoknot structure by means of NMR spectroscopy (Kolk et al., 1998a). Non-wild-type residues are given in lowercase. The structure of the 3'-hairpin, highlighted in black, was derived by NMR spectroscopy in a separate study (Kolk et al., 1998b).*

Recently, the structure of the TYMV pseudoknot has been elucidated at atomic resolution by means of high resolution NMR (Kolk *et al.*, 1998a). The system considered matches the complete T-arm and valine acceptor arm of the TYMV tRNA-like structure except for some structurally silent alterations in the stem and loop of the T-arm (Figure 3.2c). The molecule forms an elongated structure in which all three helices are stacked coaxially (Figure 3.2c). The NMR results indicate that slight torsional and swaying motions may occur between these helices on a millisecond time scale. As mentioned, the right-handedness of the resulting helix requires that loop 2 and 1 cross the minor and major groove of stem 1 and 2, respectively, and this indeed is observed. The loops are much more structured than originally envisaged on the basis of introductory NMR experiments (Puglisi *et al.*, 1990) and computer modeling based on biochemical data (Dumas *et al.*, 1987). Figure 3.3 shows the detailed structure of loop 2 crossing the minor groove of stem 1. The loop composed of three residues, U33-A35, spans a distance of ~14 Å generated by three base pairs. A35 is almost perpendicular with respect to the last base pair of stem 2 and is hydrogen bonded to G30 and G31 which are part of stem 1. C34 stacks on A35 but seems more flexible; its amino protons are within hydrogen bonded distance of the 2'-hydroxyl group of G31 but only part of the time. On the other hand, loop 1 crossing the major groove of stem 2 consists of four nucleotides, C21–U24. Interestingly, U24 and U23 continue to stack on stem 2 in an A-type helical fashion, thereby effectively increasing stem 2 to a 'seven base pair helix', upon which, after a sharp turn, the created major groove is spanned by the two remaining residues (see Figure 3.4). At this point it is important to realize that continued stacking of U23 and U24 in fact reduces the distance to be spanned over the major groove to about 11 Å which can be achieved by two nucleotides.

The junction between its two stems is an important characteristic of a pseudoknot. In the TYMV molecule, the aforementioned interactions of G30 and G31 with A35 deeply anchor the latter in the minor groove of stem 1, a surprising result in view of the presumed shallowness of the minor groove and the sensitivity of A35 toward the single-strand-specific chemicals diethyl pyrocarbonate and sodium bisulfate (Van Belkum *et al.*, 1989). A view along the helix axis shows how this is brought about. The adjacent base pairs A29–U36 and C20–G30 of stem 2 and stem 1, respectively, exhibit a helical twist angle of 52° which represents an increase of 18° with respect to the standard A-type helix (Figure 3.5). As a result, the minor groove opens up so that the interaction of A35 with G30 and G31 can

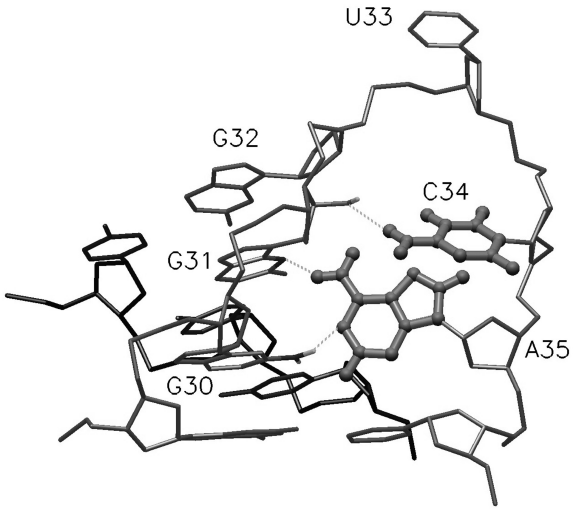

Figure 3.3 *Details of part of the TYMV pseudoknot structure showing the hydrogen bonds (dashed lines) found between loop2 and stem 1 (see text).*

be accommodated easily, which presumably contributes significantly to the stability of the junction.

The formation of this part of the pseudoknot can also be looked upon from a different perspective namely the interaction of the 3'-hairpin loop (see Figure 3.2c) with the 5'-single-stranded tail of this hairpin. This hairpin loop consists of six nucleotides, residues G30-A35. The isolated 3'-hairpin loop could therefore form a sheared GA base pair as is observed in GNRA tetra-loops (Heus and Pardi, 1991) and in pentaloops with a GNRA fold in *boxB* RNA (Cai *et al.*, 1998, Legault *et al.*, 1998). It has been suggested that any loop having a GNR$(N)_x$A (x = 1,2,3,…) consensus sequence may form such a GNRA fold (Legault *et al.*, 1998, Abramotvitz and Pyle, 1997). The 3'-hairpin loop belongs to this class and one might think that formation of a GA sheared base pair might contribute significantly to the stability of this loop as it does in the GNRA folds. This would take away the free energy gain surmised to have been obtained upon adoption of the aforementioned particular configuration of and around the junction. If during the folding of the pseudoknot first

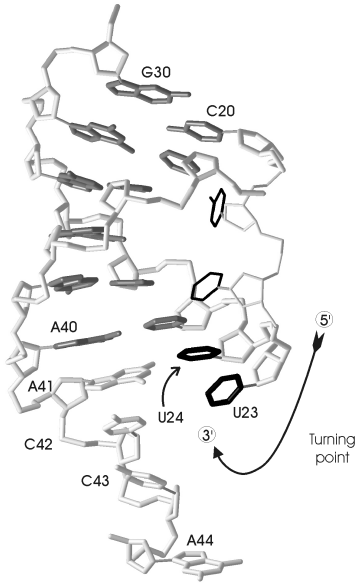

Figure 3.4 *View into the major groove of stem 2 (bases indicated in grey) of the TYMV pseudoknot structure showing the distinct turn in loop 1 (bases indicated in black). U24 and U23 are not involved in base pairing interactions with the opposite bases of A41 and C42. Nucleotides C21 and U22, spanning the major groove, are drawn in thin lines. The curved arrow denotes the direction of the RNA chain.*

a 3'-hairpin loop is formed with a sheared GA base pair or perhaps additional GC pair interactions, a significant rearrangement of the loop structure is required which may be energetically costly or considerably slow down the kinetics of pseudoknot formation.

Nuclear magnetic resonance studies of the isolated 3'-hairpin showed, however, that the loop structure closely resembles that in the pseudoknot (Kolk *et al.*, 1998b). Its conformation is such that the sequence C18-C19-C20 in the pseudoknot could easily pair to G30-G31-G32 in the loop. Thus, in the free hairpin G30 and G31 continue to stack on A29 of stem 2, but they are turned already in the direction of the major groove to facilitate base pairing with C20 and G19. Furthermore, given the S-type sugar conformation of G32 this base can easily be made to do so. This configuration also allows A35 to turn into the loop and form a base triple with G30 and G31 strongly resembling that in the pseudoknot. These

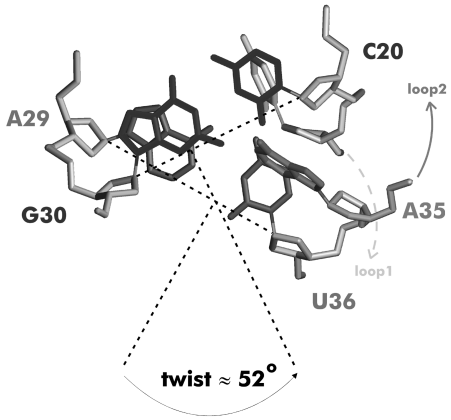

Figure 3.5 *View along the helical axis of the bases at the junction of the TYMV pseudoknot. A29-U36 is the first base pair of stem 2 and G30-C20 the last base pair of stem 1 (see also Figure 3.2c). The helical twist of the G_{30}-C_{20} and A_{29}-U_{36} base pairs is ~52°. The direction of the loops is denoted by arrows.*

results indicate that the 3'-hairpinloop is preformed to facilitate the overall folding of the pseudoknot.

Obviously, the information determining the loop structure formation is embedded in the sequence and in the present case it is clear that it dictates more than just base-pair formation and folding according to one of the major families of loop structures in particular the GNRA fold. For the 3'-hairpin the mechanism responsible for not folding into this conventional motif may be related to A29 (or more generally a purine in that position) preceding G30 (Serra *et al*, 1993, Hilbers *et al.*, 1994).

Pseudoknots involved in frameshifting

The mouse mammary tumor virus (MMTV) pseudoknot

The mechanism of frame shifting is not well understood, but recently, in a series of NMR experiments, Tinoco and collaborators have explored structural pseudoknot features that may be important in the process. The studies were conducted on pseudoknots derived

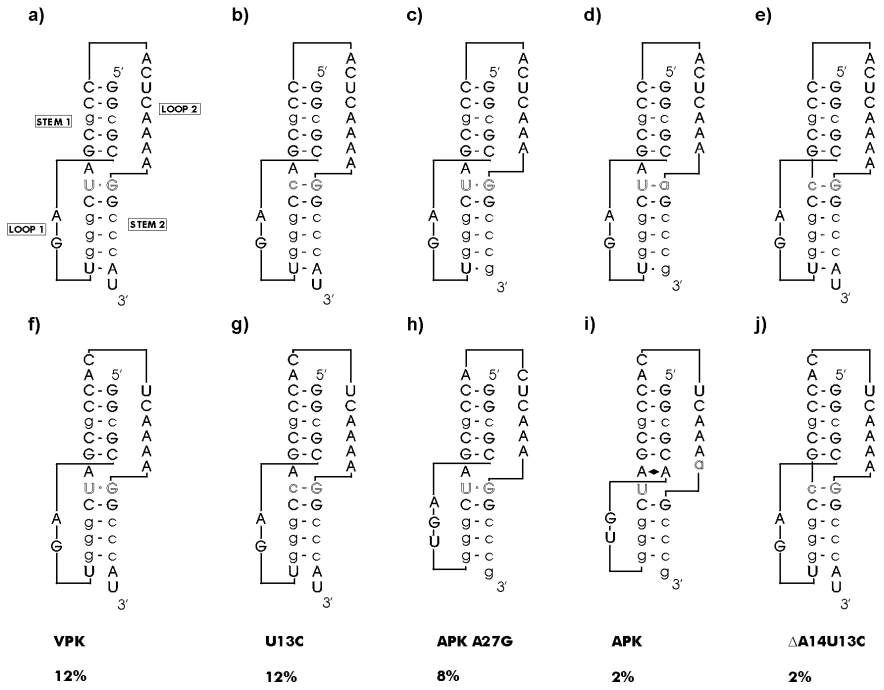

Figure 3.6 *Secondary structure models of MMTV-derived pseudoknots predicted from their nucleotide sequence (a-e) and the corresponding secondary structures derived from NMR structure studies (f-j). Lower case letters indicate deviations from the wild-type MMTV pseudoknot (see text). Names and percentages for the frameshifting efficiency of the different pseudoknots are given below each pair of structures*

from mouse mammary tumor virus (MMTV). Many retroviruses, including MMTV, use a –1 frameshift translation mechanism for the precise control of the production of the relative amounts of essential viral proteins such as the structural Gag and the enzymatic Pro/Pol proteins. The -1 frameshifting in the *gag/pro* gene overlapping region of the MMTV mRNA is promoted by at least two signals one of which is a pseudoknot (Chamorro *et al.*, 1992). The secondary structures for different MMTV-derived pseudoknots are presented in Figure 3.6. The upper panel presents the secondary structures expected on the basis of the nucleotide sequence, the lower panel those found in the derived NMR structures. The molecule, designated VPK and depicted in Figure 3.6a and 3.6f, is representative of the

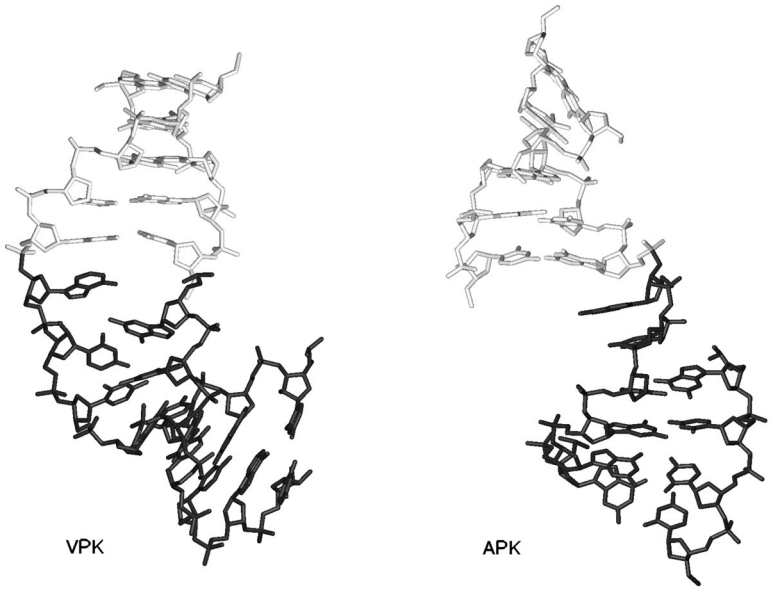

Figure 3.7 *Comparison of the mutual orientation of stem 1 (light) and stem 2 (dark) in the VPK (highly efficient frameshifting) and APK (less-efficient frameshifting) pseudoknots.*

pseudoknot in MMTV. Its sequence differs only slightly from that of the wild-type pseudoknot in that four of the G.C base pairs were switched to C.G pairs (lower case symbols) in VPK. The wild-type and the VPK sequence were shown equally efficient in promoting ribosomal frameshifting (Chen *et al.*, 1995, Shen and Tinoco, 1995). The other pseudoknots in Figure 3.6 are variants of the VPK molecule but show strong variations in frameshifting ability. A typical aspect of the expected secondary structures in Figure 3.6a-3.6d is the intercalated adenine residue in between the two projected helices. This prevents a direct stacking of the two helices upon one another in contrast to what is seen in most of the other pseudoknots discussed in this review. In particular, in the pseudoknot from Beet Western Yellows Virus (BWYV), which also promotes ribosomal frameshifting, the stems are stacked (vide infra).

The experimentally determined structures of the MMTV-derived pseudoknots that induce frameshifting indeed have this common adenine intercalated between the two

helices in their predicted secondary structure counterparts (compare Figures 3.6f-h). The situation differs for the molecules that give rise to negligible frameshifting. There the introduced mutations give rise to important changes in the structure around the junction (Figures 3.6i and j). It is of some interest to consider the different mutations involved in somewhat more detail. The first mutant to examine is designated APK for which the frameshift efficiency is reduced to 2% (Figures 3.6d and i). The fundamental difference is the replacement of the U.G pair on top of stem 2 in pseudoknot VPK by a UA pair, through deletion of a guanine residue, in APK (Figure 3.6d). In fact, in the latter molecule this base pair is not formed; instead the adenine of this base pair is incorporated in loop 2 and the uridine remains unpaired. Furthermore, the adenine of loop 1 is moved to the bottom of stem 1 forming a non-canonical base pair with the adenine in the opposite strand (Figure 3.6i). As a result the mutual orientation of stems 1 and 2 in pseudoknot VPK and APK are quite different (cf. Figure 3.7). In VPK the bend at the stem-stem junction has been estimated to amount to about 70º from a linear structure at the junction of two stems (Shen and Tinoco, 1995). In the APK molecule the two stems are strongly twisted and bent relative to each other and stacking has been completely lost, even to such an extent that loop 2 no longer crosses the minor groove of stem 1 (Kang *et al.*, 1996). One might argue that the shortening, by one adenine residue, of loop 2 in APK compared to VPK might be responsible for the drastic structural changes between the two molecules, but this is refuted by the results obtained for the mutant APK A27G (Kang and Tinoco, 1997). In this molecule the U.G base pair on top of stem 2 is restored to the situation in VPK, but loop 2 of APK is retained. The conformation of APK A27G is very similar to that of VPK and U13C (vide infra) and its frame-shifting efficiency has gone up again to 8%. In the U13C pseudoknot the U.G pair on top of stem 2 in VPK has been replaced by a C.G pair (compare Figures 3.6a and b). This point mutation leaves the structure of U13C virtually unaltered relative to that of VPK and retains the frame-shifting efficiency at 12%. These results suggested that the bent conformation induced by the hinge, the adenine residue between stem 1 and 2, is an important determinant for the frameshifting ability of the pseudoknots. Indeed, for the mutant A14U13C in which this adenine had been deleted (compare with the parent compound U13C in Figure 3.6b), a pseudoknot is obtained with stem 1 and 2 coaxially stacked on each other (Chen *et al.*, 1996) and concomitantly its frameshifting ability reduced to 2%.

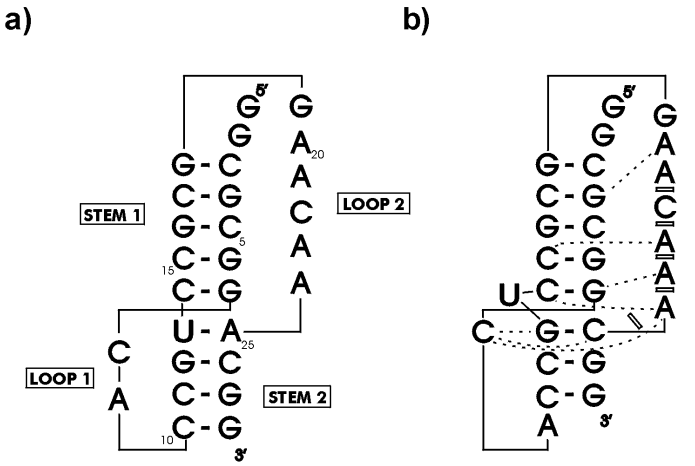

Figure 3.8 *a) Secondary structure predicted from the nucleotide sequence and b) the actual secondary structure as found in the X-ray structure of Beet Western Yellows Virus (BWYV). Stacking interactions in loop 2 are represented by open bars; the dotted lines denote base-to-base hydrogen bond interactions of loop residues with stem 1.*

The beet western yellows virus (BWYV) pseudoknot

Pseudoknots with such low frame-shifting efficiencies should not be dismissed too easily as being biologically unimportant. A case in point is the BWYV pseudoknot; *in vitro* and *in vivo* experiments by Pleij and collaborators have shown that it has a frame-shifting efficiency of 1% and that the frameshifting in BWYV is pseudoknot dependent (Garcia *et al.*, 1993). It is noted in passing that the normal level of frameshifting is of the order of 1 in 10^4. Construction of its secondary structure (see Figure 3.8a) on the basis of its sequence suggests that it adopts a classical pseudoknot structure in which the two stems may stack coaxially. Rather surprisingly, the recently determined crystal structure (Su *et al.*, 1999) indeed shows classical pseudoknot formation, but with a somewhat different secondary structure in which uridine13 of base pair U13A25 (Figure 3.8a) is bulged out and adenine 25 is added to loop 2 (Figure 3.8b). This unpairing has also been observed in other pseudoknots (cf. Figures 3.6d and 3.6i). As a result base pair G7C14 of stem 1 and G12C26 of stem 2 form the junction; the helical twist between these base pairs is large, i.e. close to

90°. This is much larger than in the classical TYMV pseudoknot. As a result, at the junction in the BWYV pseudoknot the minor groove has opened to such an extent that the adenines 24 and 25 of loop 2 are incorporated in a base quadruple through hydrogen bonding to G7 and C14 respectively (not to be confused with the quadruple G12C26C8A25 (Su *et al.*, 1999). This situation is quite distinct from the TYMV pseudoknot where the opening of the minor groove to about 52° allows a different base triple formation of A35 and G31 and G32, but base A35 has to adopt a perpendicular orientation with respect to the last base pair of stem 2 (see Figure 3.5). In the BWYV pseudoknot the mentioned interactions and the involvement of the adenine residues of loop 2 in interactions with the base pairs in the minor groove of stem 1 distort this stem with respect to the A-form helical geometry, but render the molecule quite compact. Adenine 9 of loop 1 in Figure 3.8a stacks on stem 2 in effect making it a four-base paired stem which is spanned over the major groove by one cytosine (Figure 3.8b). This is facilitated by the aforementioned large helical twist between the base pairs at the junction, by the difference of stem 2 from the standard A-form duplex RNA, and by the tilt between the two stems of about 25°.

The much smaller bending in the BWYV pseudoknot and its low frameshifting efficiency, compared to, e.g., the VPK pseudoknot discussed above, may concur with conclusions reached by Tinoco and collaborators regarding the influence of the pseudoknot structure in particular the effect of bending on the frameshifting efficiency (vide supra). In this respect it is interesting to mention the recent results obtained by Steitz and coworkers (Jaeger *et al.*, 1998). They reported the crystal structure of a complex formed between HIV-1 reverse transcriptase and an RNA pseudoknot. The latter was selected on the basis of a search for high affinity RNA ligands for the reverse transcriptase by means of the SELEX procedure (Tuerk and Gold, 1990). This pseudoknot, without a base intercalated between its two stems as is the case in the BWYV pseudoknot, turns out to be strongly bent in the protein-RNA complex. Thus, it may well be that in the latter case the protein, and for the low-efficiency frameshifting pseudoknots the ribosome, induces strong bending which is absent or much less in the free pseudoknot. This may be aided by the structural flexibility of the pseudoknots, which was shown by means of NMR to exist in the TYMV molecule (Kolk *et al.*, 1998a).

The detailed structural picture now available for the BWYV pseudoknot (Su *et al.*, 1999) shows a number of new structural motifs. We explicitly want to mention the

interactions of the adenine residues of loop 2 in the minor groove of stem 1 that give rise to new triplex motifs and the formation of the base quadruples at the junction. It is expected that these motifs will be found in other RNA molecules as well (vide infra).

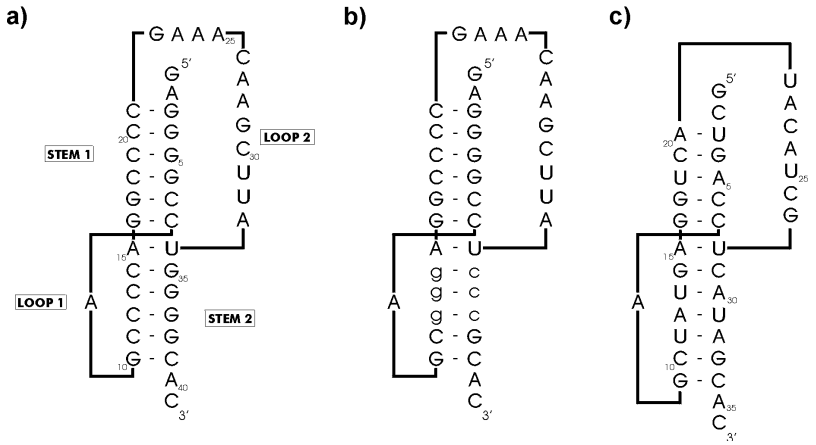

Figure 3.9 *Secondary structures of the pseudoknots from a) SRV-1, b) a SRV-1 mutant and c) gene 32 RNA of bacteriophage T2, which all belong to the group of CPK1 pseudoknots which are characterized by one nucleotide in loop1. The lower case letters in mutant SRV-1 indicate the differences with respect to wild-type SRV-1.*

The simian retrovirus type-1 (SRV-1) pseudoknot

Frameshifting pseudoknots belonging to the so-called class CPK1 (Common Pseudoknot Motif 1) have been suggested to behave (somewhat) differently from those discussed above. This family of molecules was proposed by Hoffman and collaborators (Du *et al.*, 1996) as a class of structurally related RNA pseudoknots. Important members, among others, are the autoregulatory pseudoknots within the gene-32 mRNA of bacteriophages T2 and T4 and the frameshifting SRV-1 (Simian Retrovirus type-1) pseudoknot. The secondary structures, expected for these molecules, exhibit some characteristic features. Stem 2 is composed of 6 to 7 base pairs and loop 1, which spans the major groove of stem 2, consists of a single adenosine. Furthermore, at the pseudoknot junction stem 2 ends with an AU base pair (see Figure 3.9). NMR studies of the bacteriophage T2 pseudoknot have

indicated that the two stems are stacking coaxially and that the base pair at the junction (A15-U28 in Figure 3.9c) is formed indeed (Du *et al.*, 1996). In view of the expected, common secondary structures (Figure 3.9), a similar three-dimensional structure was foreseen for the SRV-1 pseudoknot. This was investigated by subsequent NMR studies on the SRV-1 and a functional SRV-1 mutant pseudoknot. From the experiments it was concluded that these two molecules form classic H-type pseudoknots with two A-form helical stems of six base pairs and that loop 1, consisting of one nucleotide, spans the major groove of stem 2. Furthermore, the imino-proton spectra indicate that the AU base pair at the junction (A15-U43, Figure 3.9a and b) is formed. In other words, the results suggest that in these pseudoknots the two stems are coaxially stacked as in the pseudoknot from bacteriophage T2 (Du *et al.*, 1997). However, both pseudoknots had been shown to have about the same level of frame-shifting efficiency (Ten Dam *et al.*, 1995), i.e. ~ 23% and this observation and the aforementioned structural results seem at odds with the findings of Tinoco and collaborators. Two remarks have to be made at this point. First, the NMR studies conducted so far did not yet yield high-resolution structures nor has the internal flexibility of the different pseudoknots been studied in depth. For instance, the exchange rates of the iminoprotons of the AU base pairs at the junction of the frame shifting and the T2 pseudoknots appear to differ, suggesting that the frame-shifting pseudoknots are somewhat more flexible. This may have to do with the structure of stem 2. Loop 1, consisting of one adenosine, spans the major groove of stem 2 thereby crossing seven base pairs in the T2 and six base pairs in the SRV-1 pseudoknot. To what extent and how this may influence frameshifting remains to be established. However, mutational studies of the wild-type SRV-1 pseudoknot conducted by Sung and Kang (1998), in which the AU pair at the junction (A15-U43, Figure 3.9a) was replaced by either a GC pair or AA, AG, AC, GA and GG mismatches, showed that frameshifting efficiencies are similar to that in the wild-type pseudoknot. This suggested to the authors that the presence of the AU base pair between stems 1 and 2 is not essential for efficient frameshifting and that coaxial stacking is probably absent in the wild-type SRV-1 pseudoknot.

Message switching

The process of frameshifting in which pseudoknots are involved has been likened to message switching (Jentsch, 1996). In the latter process, a mRNA lacking a stopcodon is

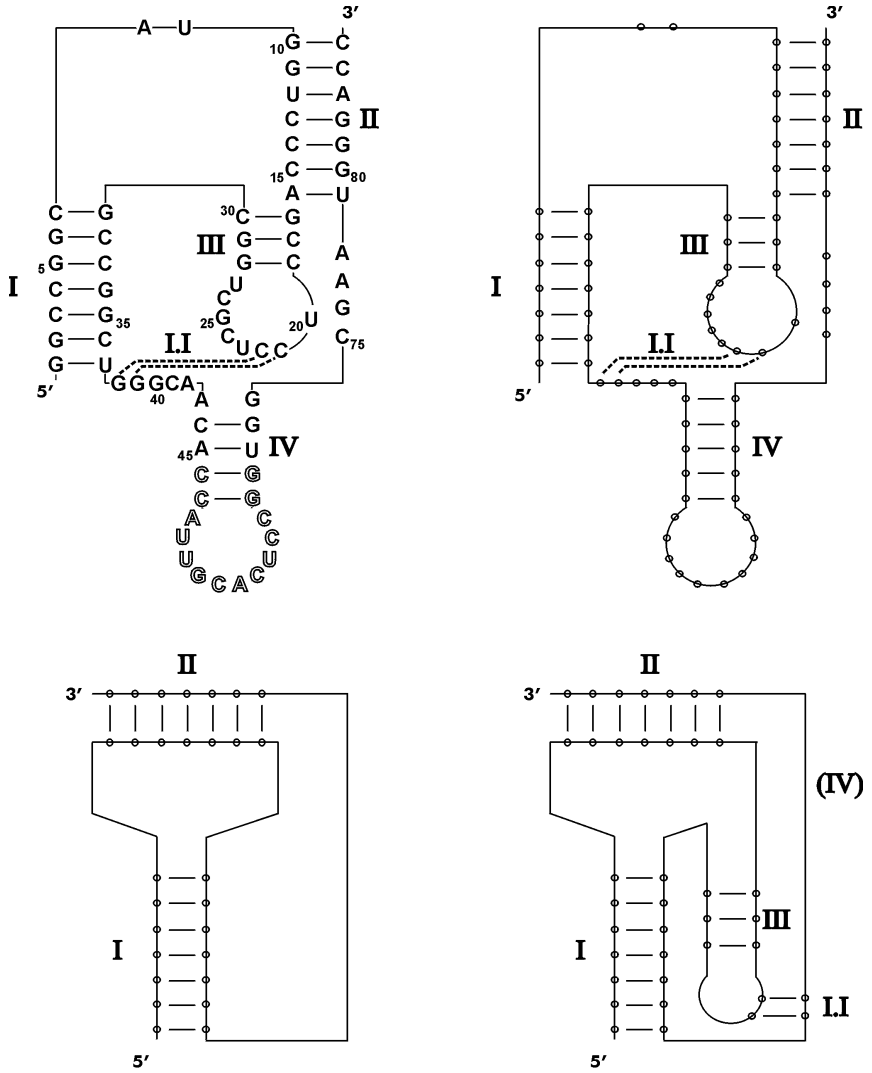

Figure 3.10 *a) Secondary structure and b) schematic representation of the hepatitis delta virus (HDV) ribozyme. The dotted lines indicate the base pairing in helix I.I which consists of two G-C base pairs. The open characters represent the binding domain of the U1A protein that was co-crystallized with the RNA. c-d) Schematic representations of the formation of the "nested" pseudoknot. For the sake of simplicity, stem-loop IV has not been drawn but its position in the sequence is marked by (IV).*

released from the ribosome which is then able to switch to the reading frame of a second RNA, the so-called tmRNA (10Sa RNA) molecule (Keiler *et al.*, 1996). Apart from a coding region this molecule also contains a tRNA-like structure, which can be charged with alanine, in addition to several pseudoknotted parts (Felden *et al.*, 1997). In either case, in frameshifting as well as in message switching, the pseudoknots may be crucial in positioning the tRNA or tRNA-like structure with respect to the mRNA or the tmRNA coding sequence, enabling the progress of these two processes. Indeed, recent experiments have indicated that a pseudoknot in *E. coli* tmRNA is involved in the *trans*-translation process (Nameki *et al.*, 1999). Of course, there is an important difference between frame shifting and message switching. Frameshifting works in *cis*, while message switching works in *trans* because the new reading frame is provided by a second molecule.

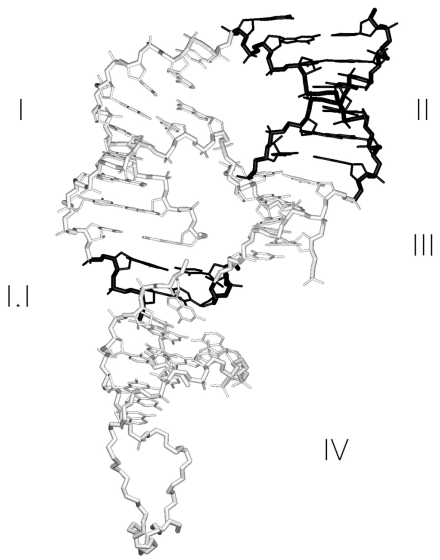

Figure 3.11 *View of the interaction scheme of the five helical regions in the HDV ribozyme. Stem I and stem II are part of one pseudoknot, stem III and stem I.I are part of the second pseudoknot. In contrast to the folding pattern of the classical pseudoknot in the HDV ribozyme stem I of the first pseudoknot is stacked on stem I.I of the second pseudoknot and stem II of the first pseudoknot is stack on stem III of the second pseudoknot.*

The hepatitis delta virus ribozyme

The structure of the most complicated pseudoknotted structure elucidated thus far is the hepatitis delta virus (HDV) ribozyme (Ferré-D'Amaré *et al.*, 1998). The structure was solved by means of X-ray diffraction and high quality crystals were obtained in a very elegant manner by engineering a high affinity binding site for the RNA binding domain of the U1A spliceosomal protein into the RNA molecule. Co-crystallization of this protein with the RNA rendered well-ordered crystals suitable for structural studies. The RNA molecule forms a compact structure containing double-helical segments which compose the scaffold for the formation of a nested double pseudoknot. How this is achieved is schematized in Figure 3.10. For comparison we consider the secondary structure originally proposed by Perrotta and Been (1991) with the addition of the helical segment, stem I.I, found in the crystal structure (Figures 3.10a and b). The first pseudoknot can be constructed from the hairpin, formed with stem I, and the interaction between the hairpin loop and the 3'-tail forming helical segment stem II (Figure 3.10c). The original hairpin loop contains a hairpin, with stem III, of which the loop forms a two base pair helical stem I.I with the 3'tail (Figure 3.10d), giving rise to a second pseudoknot nested in the first. The stacking pattern of the helical segments in this structure is different from that in the classical pseudoknot in which, on average, the two stems stack coaxially. In the HDV-molecule the situation becomes intermingled in that stem I of the first pseudoknot stacks on stem I.I of the second and stem III of the second pseudoknot stacks on stem II of the first Figure 3.11. These two helical stacks, the first one extended with helix IV of the hairpin harboring the binding site for RNA binding domain of the U1A spliceosomal protein, are arranged side by side. In this respect the structure resembles the P4-P6 domain of the *Tetrahymena* group I intron (Cate *et al.*, 1996). The way in which in the latter molecule the helical stacks are held together is, however, different from the HDV-molecule. In the P4–P6 domain the helical stacks interact through non-canonical loop-to-helix-groove contacts, while in the HDV double nested pseudoknot the helical stacks are connected by the intervening chain elements. This gives rise to interesting novel structural motifs. An example is provided by the connection from C7 of stem I in the first and G10 of stem II in the second helical stack and by C30 of stem III in the second and G31 in the first helical stack (Figure 3.11). In this way a tight connection is formed between the two helical stacks in which stem I is anchored to and becomes positioned with respect to stem II. Closer examination shows that

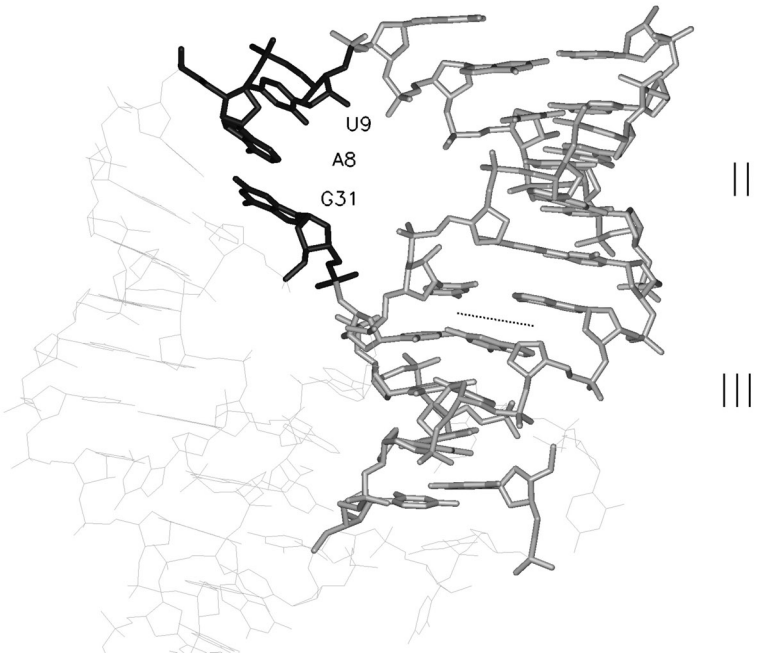

Figure 3.12 *Detail of the structure of hepatitis delta virus (HDV) ribozyme. Colored in grey are shown the coaxially stacking helices II and III with the junction between them marked with a dashed red line. The so-called pseudo-loop (colored black), consisting of G31, A8 and U9 that nicely stack on each other, spans the major groove of helix II.*

through this stacking motif not only a tight connection between the two helical stacks is formed but also a *pseudo-loop* is created that spans the major groove of stem II (Figure 3.12) in much the same way as loop 1 crosses the major groove in the classical pseudoknot (Figure 3.4). This is achieved by three residues: the first is G31, which is also involved in the formation of the terminal base pair of stem I; the second residue is A8 which stacks on G31 but resides in the chain opposite of that of G31. This is followed by U9, which stacks on A8 and is, via a sharp turn, connected to G10, which is the terminal base of stem II. Thus, the major groove of stem II, which effectively consist of seven base pairs, is crossed by a 'loop' composed of residues of different 'chains'. Recognition of the possible formation of such structural motifs may be important in mosaic modeling of RNA structures (Westhof *et al.*, 1996).

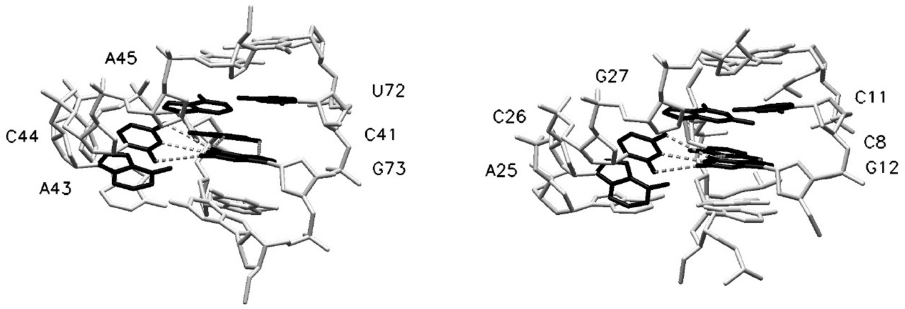

Figure 3.13 *Comparison of the CGCA base quadruple in the HDV ribozyme and the BWYV pseudoknot. There is a remarkable similarity in the configuration of all bases are annotated and shown in black.*

The position of the other end of stem I is secured via its stacking on stem I.I. G1 of the terminal wobble base pair G1U37 stacks on G38, which is part of stem I.I, of the opposite strand. At the other side helix I.I is stabilized by helix IV. Interestingly, below the junction between stem I.I and stem IV a folding pattern is obtained which is equivalent to the quadruple G12C26 C8A25 and surrounding nucleosides in the BWYV pseudoknot. These patterns are compared Figure 3.13. It is likely that the mechanism underlying the formation of these folds is the so-called Y5 rule or a variant thereof (Roberts and Crothers, 1996).

In the active ribozyme helix I bears the substrate. Therefore, the positional and orientational fixation of this helix within the molecular frame is of extreme importance for maintaining the integrity of the active site. As discussed above the X-ray structure nicely explains by what interactions these requirements are met.

Concluding remarks

The recent elucidation of a number of high-resolution pseudoknot structures underscore the versatility and complexity of RNA folding. In the process, several new structural motifs have been uncovered and there is little doubt that more will be found when

the structure of new classes of pseudoknots will be elucidated. It has also become clear that some of these motifs are likely to be of more general significance. Examples are the interactions in the minor groove of stem 1 of the TYMV pseudoknot (Kolk *et al.*, 1998a) and of the BWYV pseudoknot (Su *et al.*, 1999) as well as the formation of the base quadruples in the BWYV pseudoknot (Su *et al.*, 1999) and the HDV ribozyme (Ferré-D'Amaré *et al.*, 1998) discussed above. The results also indicate that for further progression new structures need to become available. As described above, so far only the structure of two types of pseudoknots is available, i.e. the structure of a subclass of the classical pseudoknot and the structure of the nested double pseudoknot as found in the HDV ribozyme. It was brought up in the introduction that it is not always obvious how to define a pseudoknot fold. For the classical pseudoknot structures discussed here there is an evident matching between their secondary structure model and their three-dimensional folding. For the HDV ribozyme this correspondence is less clear if not absent. Therefore it seems preferable to classify pseudoknot folds from determined high-resolution three-dimensional structures. This is common practice in protein structural studies where folds have also been classified on the basis of the three-dimensional structure data base. Again, this means that more high-resolution structures, also of other classes of pseudoknots, need to be determined.

On the basis of the newly determined pseudoknot structures also more meaningful structure-function studies can be performed. Interesting attempts have already been made for the MMTV-derived frameshifting pseudoknots by Tinoco and collaborators indicating, that a specific bend in the pseudoknot may be an important factor for efficient frameshifting. It is not clear yet, however, to what extent this aspect is important for other frameshifting pseudoknots. Conceivably, also the flexibility and stability (Ten Dam *et al.*, 1995) of the molecules will be an important issue, as well as their interaction with proteins (Spedding *et al.*, 1993, Philippe *et al.*, 1993) or other factors. In the quest for the latter, the high-resolution studies will be a prerequisite for understanding the crucial interactions and for initiation and fine-tuning of the complementary functional studies.

Chapter 4

Heteronuclear ^{1}H-^{13}C experiments for complete assignment of adenine proton and carbon resonances in $^{13}C/^{15}N$ labeled RNA

Paul J.A. Michiels, Marco Tessari, Geerten W. Vuister, Cornelis W. Hilbers, & Hans A. Heus, submitted to J. Biomol. NMR

Abstract

We present two experiments that allow unambiguous assignment of the non-exchangeable proton and carbon resonances of the adenine base moiety. In the first proton-carbon correlation experiment magnetization is first transferred from the adenine H2 and H8 resonances to the C2 and C8 spins via one-bond HC couplings and subsequently to the carbon spins C5 and C4/C6, respectively, by using a defocusing/refocusing delay tuned to the three-bond couplings. A second experiment was designed that not only allows assignment of all proton and carbon resonances but also correlates the H8 and H2 resonances. This is achieved by adding a second relay step, which correlates both aromatic ring protons to the C4, C5 and C6 resonances. Minimal loss of magnetization was achieved by transferring magnetization as heteronuclear multiple quantum coherence making the experiments suitable for larger RNA molecules. The experiments are demonstrated on the 36-nucleotide frameshifting RNA pseudoknot of SRV-1, for which unambiguous assignment of all adenine carbon and proton resonances was obtained.

Introduction

The usage of dipolar couplings obtained by alignment of molecules in anisotropic solutions has become an important tool in the structure determination of proteins and nucleic acids. Especially, the determination of dipolar couplings between covalently bound nuclei that can be translated into the relative orientations of ^{1}H-^{13}C, ^{1}H-^{15}N, ^{13}C-^{13}C and ^{15}N-^{13}C vectors has become very important because they hold long-range angular information. In particular structure determination of RNA molecules can benefit from this novel approach, since without the availability of this type of long-range information structure calculations of RNA molecules mainly depend on local sequential restraints obtained by the traditional NOE and J-coupling data, which results in less precisely determined structures (Mollova *et al.*, 2000, Lynch and Puglisi, 2000, Trantirék *et al.*, 2000). However, collection of one-bond dipolar coupling constants depends on a prior unambiguous assignment of the relevant proton and carbon resonances. The two experiments presented here concern with the assignment of the carbon and proton resonances of adenines. The first experiment we describe can be used to quickly and efficiently correlate the proton and carbon resonances. Previously, unambiguous assignment of H2 resonances has been achieved by obtaining through-bond correlations between the H8 and the H2 proton resonances from HCCH-TOCSY (Marino *et al.*, 1994, Legault *et al.*, 1995) or HNC-TOCSY-CH (Simorre *et al.*, 1996) experiments using isotopically enriched RNA. The second experiment we describe correlates the adenine H8 and H2 resonances, using an alternative magnetization transfer pathway, which might be more useful for larger molecules than these previously published experiments.

Results

The first experiment, termed H(C)C is a variation of the LRCC experiment (Bax *et al.*, 1992). This experiment, of which the pulse sequence is shown in Figure 4.1a, correlates the adenine H2 with the C2 and C5 and the H8 resonances with the C8, C4 and C6 resonances via the one bond proton-carbon and the long range carbon-carbon scalar couplings (for an overview of the scalar coupling constants in RNA building blocks, see

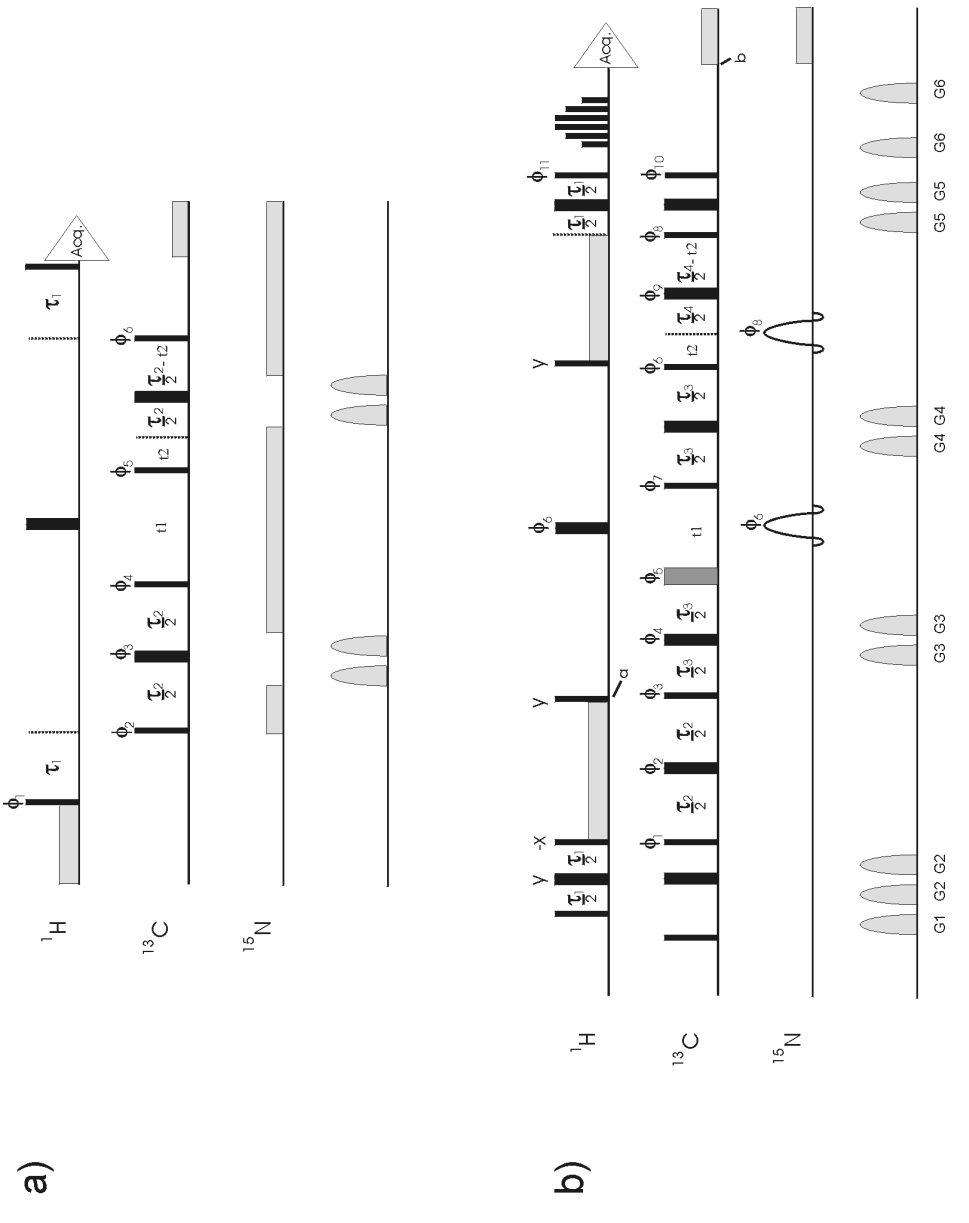
a)
b)
^{1}H
^{13}C
^{15}N
Acq.
t1
t2
G1 G2 G3 G4 G5 G6

Figure 4.1 *Pulse sequence representations of a) the 3D H(C)C and b) 3D long-range H(C)CC experiment used to assign all carbon resonances of adenine. The 90˚ and 180˚ pulses are indicated by black narrow and thick bars, respectively. All pulses have phase x unless otherwise indicated. a) In the H(C)C experiment the delay times were τ1 = 2.4 ms and τ2 = 25 ms. The ^{13}C carrier was placed in the middle of all base carbons at 147 ppm and the ^{15}N-carrier at 216 ppm. Radiofrequency field strengths of 1.36 kHz were used for ^{13}C-decoupling using GARP and 1.1 kHz for ^{15}N-decoupling using Waltz. Gradients were applied with a duration of 1 ms and with a 13 G/cm strength. The 4-step phase cycle is φ1 = y,-y; φ2 = (x,-x); φ3 = x; φ4 = y ; φ5 = y ; φ6 = x, receiver = x. In addition, φ2,φ3,φ4 and φ5, φ6 are cycled to obtain quadrature detection using States-TPPI for t1 and t2, respectively. Sweep widths of 8000, 2800 and 10000Hz were used in t1, t2 and t3, respectively, 4 scans for each FID, 40 complex points in t1, 80 complex points in t2, and 768 complex points in t3. The total experimental time was 14 hrs. b) In the 3D long-range H(C)CC experiment prior to t1 evolution a 270˚ pulse on carbon is applied, indicated as a thick grey bar. Nitrogen decoupling during t1 and t2 is performed by 1.5 ms hyperbolic secant pulses. The delay times were τ1 = 2.4 ms, τ2 = 17 ms, τ3 = 2.5 ms and τ4 = 22 ms. The ^{1}H carrier was placed midway in between the H2 and H8 resonances at 7.87ppm during the CW spinlock pulse on proton using a 654 Hz RF-field strength. The ^{13}C-carrier was placed in the center of all base carbons at 139 ppm and the ^{15}N-carrier frequency at 152 ppm in the center of all base nitrogens and was changed prior to acquisition to 195 ppm (point b) for decoupling. Radiofrequency fields for decoupling were 1.36 kHz for 13C using GARP and 1.1 kHz for ^{15}N using Waltz. Off-resonance WATERGATE (phases = x,-x,x,x,-x,x) was performed to suppress the residual water. Gradients with a duration of 1 ms were applied with the strengths; G1 = 23 G/cm, G2 = 6 G/cm, G3 = 9 G/cm, G4 = 7 G/cm, G5 = 15 G/cm and G6 = -13 G/cm. The 32-step phase cycle is φ1 = x; φ2 = 8(x,-x); φ3 = (-x,x); φ4 = x; φ5 = 2(y,-y) ; φ6 = 8(x,-x) ; φ7 = 4(y,-y) ; φ8 = x; φ9 = 4(x,-x,y,-y) ; φ10 = 16(x,-x) ; φ11 = 4(y,-y,) receiver = 2(x,-x,-x,x, -x,x,x,-x). In addition, φ1,φ2,φ3,φ4,φ5 and φ7,φ8 are cycled to obtain the States-TPPI quadrature detection for t1 and t2, respectively. Sweep widths of 5494.5, 2793.3 and 10000 Hz were used in t1, t2 and t3, respectively, 32 scans for each FID, 39 complex points in t1, 52 complex points in t2, and 512 complex points in t3. The total experimental time was 63 hrs.*

Ippel *et al.*, 1996, Wijmenga and van Buuren, 1998). The relevant pathways of magnetization transfer starting from adenine H2 or H8 are indicated in Figure 4 2a.

In our experiments we have exploited the longer relaxation times of multiple quantum coherence (Grzesiek and Bax, 1995, Marino *et al.*, 1997) to achieve minimal loss of signal intensity. First, multiple quantum $H2_yC2_y$ magnetization is created via the large $^1J_{CH}$ scalar coupling of approximately 200-216 Hz, using a defocusing delay $\tau_1 = 1/2J_{CH}$, of 2.4 ms, followed by a 90^0_x pulse on carbon. During the second delay τ_2 the multiple quantum $2H2_yC2_y$ coherence is partly converted into $4H2_yC2_xC5_z$ via the $^3J_{C2C5}$ scalar coupling. Then, by applying a ^{13}C 90^0_y pulse $4H2_yC2_xC5_z$ is converted to $4H2_yC2_zC5_x$. Starting at H8, and applying the same coherence transfer pathway, yields both in $4H8_yC8_zC4_x$ and $4H8_yC8_zC6_x$. After t_1 evolution, with frequency labeling on carbon, the coherences are refocused back to H2 and H8 protons by the reverse pathway. In order to decrease resonance overlap the experiment was recorded in a 3D fashion, by inserting a

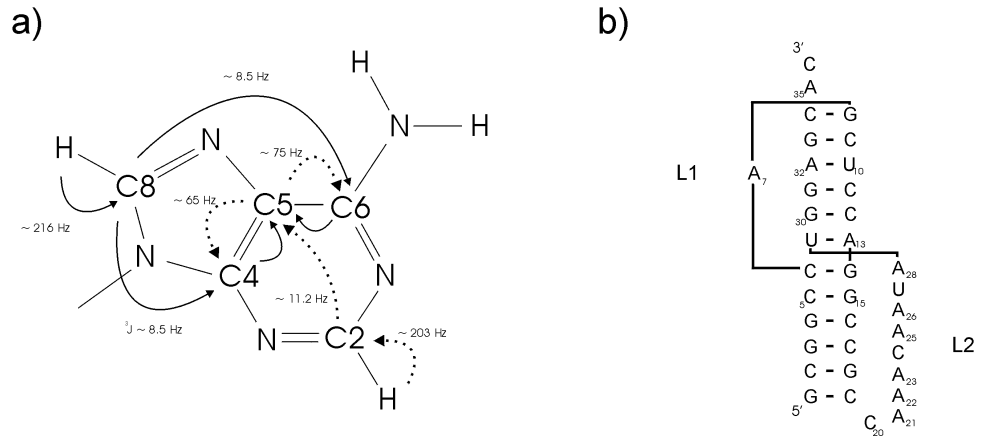

Figure 4.2 *a) Schematic representation of the adenosine base moiety. Arrows indicate the two different pathways of magnetization transfer that correlate the H2 and H8 resonances with the mutually coupled C4, C5 and C6 nuclei. b) Secondary structure model of the SRV-1 frameshifting RNA pseudoknot, used in this study. L1 and L2 indicate the two loops present in this type of pseudoknot.*

constant-time chemical shift evolution on carbon - monitoring the C2/C8 region – during the last τ_2-refocusing step. Because of its simplicity and high sensitivity the H(C)C experiment is very useful in quickly assigning all carbon resonances within the adenine aromatic ring. The experiment can, however, be deficient in assigning all carbon resonances unambiguously in the case of overlapping proton resonances and depends on prior assignment of the H8 and H2 resonances by other methods.

Therefore, a second experiment, termed 3D long range H(C)CC, was designed which extends the H2{C2, C5} correlation to C4 and C6 and the H8{C8, C4, C6} correlation to C5 by inserting an additional relay step (Figure 4.1b). A description of the coherence transfer pathway in terms of operator formalism is given in the following.

$$H2_y \xrightarrow{J_{HC}} 2H2_xC2_z \xrightarrow{90^x_C} 2H2_xC2_y \xrightarrow{J_{CC}} 4H2_xC2_xC5_z$$

$$\xrightarrow{90^x_C} 4H2_xC2_xC5_y \xrightarrow{90^y_H} 4H2_zC2_xC5_y \xrightarrow{J_{CC}} 4H2_zC2_xC5_y + 8H2_zC2_xC5_xC4_z +$$

$$8H2_zC2_xC5_xC6_z + 16H2_zC2_xC5_yC4_zC6_z \xrightarrow{90^y_C} 4H2_zC2_zC5_y - 8H2_zC2_zC5_zC4_x -$$

$$8H2_zC2_zC5_z\,C6_x + 16H2_zC2_zC5_yC4_xC6_x$$

$$H8_y \xrightarrow{J_{HC}} 2H8_xC8_z \xrightarrow{90^x_C} 2H8_xC8_y \xrightarrow{J_{CC}} 4H8_xC8_xC4_z + 4H8_xC8_xC6_z +$$

$$8H8_xC8_yC4_zC6_z \xrightarrow{90^x_C} 4H8_xC8_xC4_y + 4H8_xC8_xC6_y - 8H8_xC8_zC4_yC6_y \xrightarrow{90^y_H}$$

$$4H8_zC8_xC4_y + 4H8_zC8_xC6_y - 8H8_zC8_zC4_yC6_y \xrightarrow{J_{CC}} 4H8_zC8_xC4_y + 4H8_zC8_xC6_y +$$

$$8H8_zC8_xC4_xC5_z + 8H8_zC8_xC6_xC5_z - 8H8_zC8_zC4_yC6_y - 16H8_zC8_zC4_yC6_xC5_z -$$

$$16H8_zC8_yC6_yC4_xC5_z \xrightarrow{90^y_C} 4H8_zC8_zC4_y + 4H8_zC8_zC6_y - 8H8_zC8_zC4_zC5_x -$$

$$8H8_zC8_zC6_zC5_x + 8H8_zC8_xC4_yC6_y - 16H8_zC8_xC4_yC6_zC5_x - 16H8_zC8_yC6_yC4_zC5_x$$

In this experiment, starting from H2, first multiple quantum $2H2_xC2_y$ magnetization is created during the τ_1 delay of 2.4 ms and subsequent 90^0_x pulse on carbon. Multiple quantum $2H2_xC2_y$ coherence is partly converted to $4H2_xC2_xC5_z$ during the delay τ_2 via the $^3J_{C2C5}$ scalar coupling. The CW spinlock on proton during τ_2 evolution is used to effectively suppress modulation on carbon due to long range scalar couplings of the H2 to C4 and C6. At time point **a** ^1H 90^0_y and ^{13}C 90^0_x pulses convert $4H2_xC2_xC5_z$ to $4H2_zC2_xC5_y$. In the

following period τ_3 the $4H2_zC2_xC5_y$ coherence is converted into $8H2_zC2_xC5_xC4_z$ and $8H2_zC2_xC5_xC6_z$ via $^1J_{C5C4,6}$ scalar couplings.

In the parallel pathway, starting at H8, multiple quantum $2H8_xC8_y$ coherence is converted to $4H8_xC8_xC4_z$ and $4H8_xC8_xC6_z$ during τ_2 and successively to $8H8_zC8_xC4_xC5_z$ and $8H8_zC8_xC6_xC5_z$ during τ_3. The duration of the defocussing delays τ_2 and τ_3 was optimized to yield maximal intensity of the H2 and H8 to C4, C5 and C6 crosspeaks, while minimizing the contribution of triple quantum coherences outlined in the product operator diagram above. Like in the H(C)C experiment, after t_1 evolution coherence is refocused back to H2 and H8 protons by the reverse pathway. Again, for resolution enhancement the long range H(C)CC experiment was recorded in a 3D fashion, by recording C2/C8 chemical shift evolution during the last refocusing period (τ_4). In the 3D long range H(C)CC experiment the C4, C5 and C6 resonances are coupled to both the H2 and H8, which provides for a means to obtain through-bond correlations between these protons.

Both pulse sequences were applied on a 36-nucleotide RNA pseudoknot of simian retrovirus type-1 (SRV-1) RNA, involved in ribosomal frameshifting (Figure 4.2b, ten Dam *et al.*, 1994), which contains ten adenine residues. Two of these are involved in Watson-Crick base pairing, one is at the 3'-dangling end and the others are present in the two loops designated as L1 and L2. The single relayed H(C)C experiment provided an almost complete assignment of all adenine carbon resonances from the nicely resolved H2{C2, C5}and H8{C8, C4, C6}crosspeaks. Figure 4.3 shows examples of the H2 and H8 connectivities of A13, involved in the A-U base pair at the junction between stems S1 and S2, and the loop L2 residues A23, A25 and A28. The H2 to C2 and C5 crosspeaks are clearly discernable in the strips taken at the appropriate H2/C2 frequencies, while the H8 to C8, C4 and C6 crosspeaks can be recognized from the H8/C8 strips. However, both the H8 and C8 resonances of A23 and A25 overlap (a strong single peak appears at these frequency coordinates in the HMQC spectrum, not shown) and show two crosspeaks to the C4 region, which cannot be assigned to either A23 or A25 using this experiment.

The 3D long range H(C)CC experiment with an additional relay step resolves this ambiguity. In the spectrum (Figure 4.4) connectivities to the C4, C5, and C6 resonances are observed for the H2 and H8 of all adenine residues. Since, contrary to the H8 resonances, the H2 resonances of A23 and A25 are resolved, the C4 resonances of these particular residues can be assigned by this experiment. Thus, as can be seen from the H2/C2 strips the

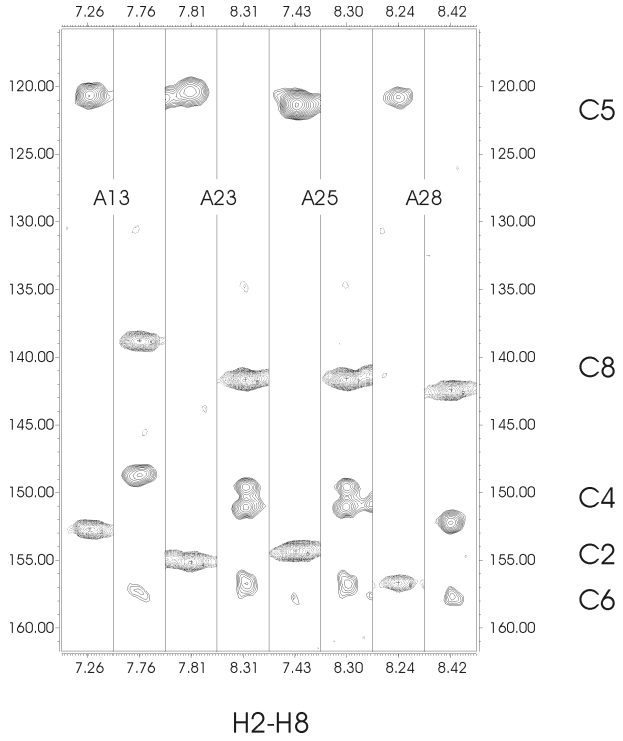

Figure 4.3 *Strips along the F2 dimension of the H(C)C spectrum showing the pairwise correlation of the H2 (left) and H8 (right) resonances with C5 and C4/C6 resonances for adenosine residues 13, 23, 25 and 28 of the SRV-1 pseudoknot. Strips were taken at the appropriate C2 or C8 plane (F2) and H2 or H8 frequency (F3).*

H8C4 cross peak at 151.1 ppm belongs to A23 and the crosspeaks at 149.6 ppm belongs to A25. Also the A23 and A25 H8-C6 correlations, which appear as a broad single crosspeak in the H8/C8 strips in the H(C)C experiment are resolved in the H2/C2 strips of the long-range H(C)CC experiment. Again, the downfield crosspeak, resonating at 157.1 ppm can be assigned to A23 H2C6 and the upfield crosspeak, resonating at 156.4 ppm to A25 H2C6.

Additionally, nearly all H2 and H8 protons can be correlated in this experiment by coupling to their common C4 and C6 partners. For instance, residues A21, A22 and A23, which have identical C4 chemical shifts, can be easily assigned via the difference in the C6

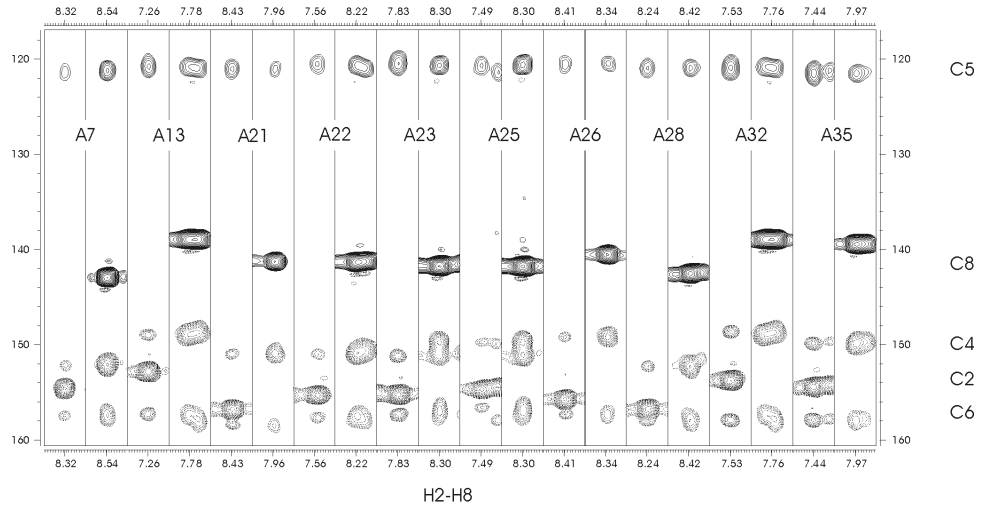

Figure 4.4 *Strips along the F_2 dimension of the 3D long-range H(C)CC spectrum showing pairwise the correlation of the proton H2 (left) and H8 (right) resonances with C4/C5/C6 for each adenosine residue of the SRV-1 pseudoknot. Strips were taken at the appropriate C2 or C8 plane (F_2) and H2 or H8 frequency (F_3).*

chemical shift. The correlation to the C5 region, albeit displaying the poorest dispersion of all carbon resonances, was essential in assigning residues A13 and A26.

Discussion

We have presented two novel experiments, which combined usage allows for a complete assignment of the non-exchangeable proton and carbon resonances of the adenines. Additionally, the second experiment we describe also correlates the H2 and H8 resonances.

NOEs involving the non-exchangeable H2 of adenines often provide valuable long-range distance information, which can be extremely useful for the structure determination of larger RNA molecules. For instance, a loop adenine residue at the junction of the turnip yellow mosaic virus (TYMV) pseudoknot adopts a very specific conformation, which could not have been revealed without the large number of NOEs between its H2 and protons of other residues at the junctional region. (Kolk *et al.*, 1998a). Likewise, in the case of the

SRV-1 pseudoknot, used in this study, a large number of long range NOEs involving adenine H2s determined a number of base triple interactions between loop L2 and stem S1 (Michiels *et al.*, 2001).

Unambiguous assignment of the adenine H2 resonances are unfortunately not straightforward because they often rely on through-space interactions. It is better to make use of the H8 proton, which is relatively easy to assign by the large number of intra-residual and sequential NOEs to the sugar protons (Wijmenga *et al.*, 1993) or by through-bond H1'-H8 connectivities using isotopically enriched RNA (Farmer *et al.*, 1993a&b, Sklenár *et al.*, 1993a&b, Tate *et al.*, 1994, Sklenár *et al.*, 1998, Riek *et al.*, 2001). Subsequently, the adenine H2 can be unambiguously assigned by obtaining throughbond H8-H2 correlations using isotopically enriched RNA (Marino *et al.*, 1994, Legault *et al.*, 1995, Simorre *et al.*, 1996).

The second, long range H(C)CC, experiment unambiguously assignes the H2 resonances by the mutual crosspeaks of H8 with C4, C5 and C6. Our approach is similar to the natural abundance HMBC experiment performed on a DNA triplex, which correlates both protons directly to their mutual $^{3}J_{HC}$-coupled C4 (van Dongen *et al.*, 1996). However, the additional use of the C5 and C6 resonances in the long-range H(C)CC experiment enlarges the chance to find a unique combination of crosspeaks.

Previously, through-bond correlations between the H8 and the H2 resonances in isotopically enriched RNA have been obtained from HCCH-TOCSY (Marino *et al.*, 1994, Legault *et al.*, 1995) or the HNC-TOCSY-CH (Simorre *et al.*, 1996) experiments. However, these experiments can be insensitive for larger molecules due to the unfavorable relaxation properties of the carbons and the long FLOPSY or DIPSI transfer times. In our experiments minimal loss of signal intensity is achieved by exploiting the longer relaxation times of multiple quantum coherence (Grzesiek and Bax, 1995, Marino *et al.*, 1997), and using a shorter transfer pathway. Additionally, the sensitivity of our experiment does not suffer from dynamics or chemical exchange of the amino protons used in the HNC-TOCSY-CH experiment. Such behavior was observed for some of the resonances of the 30 nucleotide (nt) leadzyme and the 44nt hammerhead ribozyme (Simorre *et al.*, 1996). While the sensitivity of the H(C)CC experiment was found to be comparable to the traditional 3D HCCH-TOCSY (Bax *et al.*, 1990) for the 36-mer pseudoknot, the H2 and H8 correlation

experiment presented here using a shorter transfer pathway and multiple quantum coherence, might be more suitable for larger molecules.

In conclusion, the two novel experiments we describe unambiguously assign all proton and carbon resonances of adenines. This is a necessary first step in collecting dipolar coupling and distance restraints and as such is very useful in obtaining accurate and precise three-dimensional structures of RNA molecules by means of NMR spectroscopy.

Materials and methods

A 1.8 mM $^{13}C/^{15}N$ adenine labeled sample of the simian retrovirus type-1 36-mer RNA pseudoknot was prepared enzymatically through in vitro transcription using T7 RNA polymerase (Milligan *et al.*, 1987). The product was purified using preparative polyacrylamide gel electrophoresis and electroelution. Purified RNA was extensively dialyzed against 10 mM K_2HPO_4/KH_2PO_4, 100 mM NaCl buffer, pH 6.4 and concentrated to 270 µl using a Centricon microconcentrator. Finally the sample was lyophilized and dissolved in 270 µl 99.99% D_2O. The spectra were recorded at 30°C on a Varian Inova 500 MHz spectrometer. Spectra were processed with NMRpipe (Delaglio *et al.*, 1995) and analyzed with the Xeasy program (Bartels *et al.*, 1995). All carbon chemical shifts assigned using both experiments are listed in Table A.III of the Appendices chapter in this thesis.

Chapter 5

Solution structure of the pseudoknot of SRV-1 RNA, involved in ribosomal frameshifting

Paul J.A. Michiels, Alexandra Versleijen, Paul W. Verlaan[Y], Cornelis W.A. Pleij[Y], Cornelis W. Hilbers and Hans A. Heus (2001), *J. Mol. Biol.*, **310**, 1107-1122.

Abstract

RNA pseudoknots play important roles in many biological processes. In the simian retrovirus type-1 (SRV-1) a pseudoknot together with a heptanucleotide slippery sequence are responsible for programmed ribosomal frameshifting, a translational recoding mechanism used to control expression of the Gag-Pol polyprotein from overlapping *gag* and *pol* open reading frames. Here we present the three-dimensional structure of the SRV-1 pseudoknot determined by NMR. The structure has a classical H-type fold and forms a triple helix by interactions between loop 2 and the minor groove of stem 1 involving base-base and base-sugar interactions and a ribose zipper motif, not identified in pseudoknots so far. Further stabilization is provided by a stack of five adenines and a uracil in loop 2, enforcing a cytidine to bulge. The two stems of the pseudoknot stack upon each other demonstrating that a pseudoknot without an intercalated base at the junction can induce efficient frameshifting. Results of mutagenesis data are explained in context with the present three-dimensional structure. The two base pairs at the junction of stem 1 and 2 have a helical twist of approximately 49º allowing proper alignment and close approach of the three different strands at the junction. In addition to the overwound junction the structure is somewhat kinked between stem 1 and 2, assisting the single adenosine in spanning the major groove of stem 2. Geometrical models are presented that reveal the importance of the magnitude of the helical twist at the junction in determining the overall architecture of classical pseudoknots, in particular related to the opening of the minor groove of stem 1 and the orientation of stem 2, which determines the number of loop 1 nucleotides that span its major groove.

Introduction

Simian retrovirus type-1 (Power *et al.*, 1986) uses programmed –1 ribosomal frameshifting as a mechanism for regulating translation of its polycistronic mRNA. The frameshift enables the ribosome to skip the stop codon in the 0 frame and to proceed protein synthesis until the first stop codon in the -1 frame is reached. After synthesis the polyprotein is processed by proteolysis resulting in the functional proteins pro, gag and pol. Programmed ribosomal frameshifting has not only been observed for retroviruses, but also in coronaviruses (Brierley *et al.*, 1989), plant viruses (Garcia *et al.*, 1993), yeast (Tzeng *et al.*, 1992, Dinman and Wickner, 1992), and bacterial systems (Larsen *et al.*, 1994, Gesteland and Atkins, 1996).

Two elements in the mRNA are responsible for efficient ribosomal frameshifting, i) a heptanucleotide sequence and ii) a secondary structural element 6-8 nucleotides downstream of this sequence, which in almost all cases is a pseudoknot. The heptanucleotide or so-called shifty sequence has an X XXY YYZ consensus, where X can be any base, Y is A or U and Z is A, U or C. It is thought that when this sequence enters the A and P site of the ribosome, the pseudoknot structure causes pausing of the ribosome giving time to shift to the -1 reading frame. Although in most cases a pseudoknot is found to be responsible for ribosomal frameshifting, also a hairpin structure in HIV (Parkin *et al.*, 1992) and Escherichia coli dnaX (Tsuchihashi, 1991, Larsen *et al.*, 1997) and a three-way junction in IS911 (Rettberg *et al.*, 1999) have been shown to be functionally active. Whether the pseudoknot structure acts as an independent unit, regulating the ribosomal frameshift by its structure or stability, or acts in concert with auxiliary proteins is still not understood. The importance of its structural identity is evident, however, since pseudoknots can for instance not be replaced by a functionally different pseudoknot (Kang *et al.*, 1996) or replaced by a hairpin with a similar base pair composition (Somogyi *et al.*, 1993).

The pseudoknot of SRV-1 RNA is of the classical H-type (Pleij *et al.*, 1985, ten Dam *et al.*, 1992, Hilbers *et al.*, 1998), with two stacked stems S1 and S2, connected by two loops L1 and L2. Mutational analysis of the pseudoknot in the overlapping region of the *gag* and *pol* genes of SRV-1 showed that the lengths of L1 and L2 and composition of the base pairs in S1 are important for its function (ten Dam *et al.*, 1994, 1995). In the SRV-1 (ten Dam *et al.*, 1995) and other frameshifting pseudoknots (Liphardt *et al.*, 1999) changing L2 nucleotides close to the junction can have drastic effects on frameshift

efficiency. These results, together with recently solved structures of other H-type pseudoknots (Kolk et al., 1998a; Su et al., 1999), suggested a prominent structural role for L2 having various tertiary interactions at the junction.

Here we present the high resolution NMR structure of the RNA pseudoknot present in the overlapping region of the *gag* and *pol* genes of SRV-1. The structure is further characterized by a highly organized L2-S1 interface, which forms a triple helix by various tertiary interactions and extensive stacking. Base pairing at the junction of the two stems is maintained, which shows that a pseudoknot without an intercalated base at the junction can induce a frameshift. The structure explains nearly all results of mutational studies obtained so far, and the architecture of the structure helps in understanding how a pseudoknot promotes frameshifting.

Results

Design of an SRV-1 pseudoknot sequence optimal for NMR

An optimal SRV-1 pseudoknot sequence was designed for the NMR experiments by introducing a few mutations, while maintaining its frameshift efficiency (Figure 5.1). i) The third G-C base pair in S2 was changed into an A-U base pair. This enabled us to distinguish the two stems - which are both very G-C rich - by NMR spectroscopy, and avoided the possibility of rearrangement of the pseudoknot into an alternative structure with two hairpins. ii) The three GCU residues in L2 were deleted. Frameshift efficiencies for mutants i) and ii) (ten Dam *et al.*, 1994, ten Dam *et al.*, 1995) were found to be higher compared to the wild type level (23%, Figure 5.1a), amounting to 25% and 30%, respectively (Figure 5.1b,c).

However, pseudoknots with a guanosine at position 20 showed a tendency to form multimers. This was perceived from a concentration-dependent behaviour of the imino proton resonances (between 10 and 300 μM). We believe that the concentration dependence originated from formation of dimer pseudoknot structures in which the G20-A21' and A21-G20' form an intermolecular sheared tandem G-A base pair motif at the dimer interface (Santa Lucia and Turner, 1993; Heus *et al.*, 1997). Mutational studies (ten Dam *et al.*, 1995) showed that changing 10 nucleotides at the 5'-side of L2 yields a

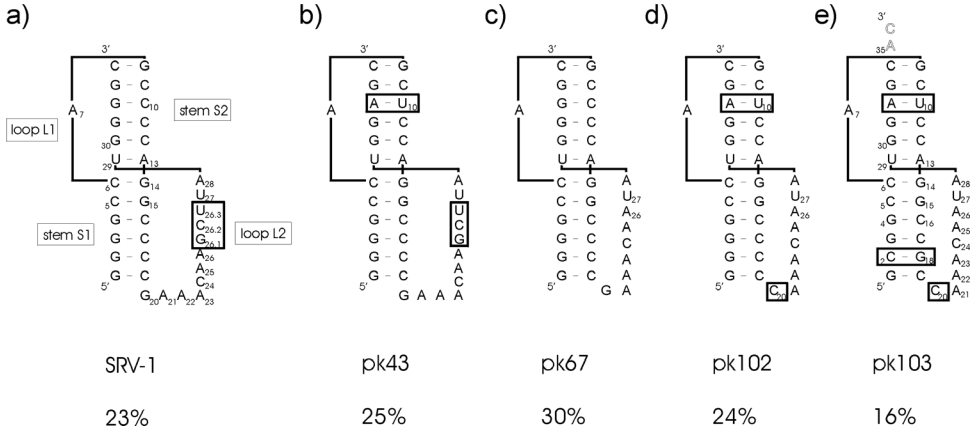

Figure 5.1 *Secondary structure representation of the SRV-1 wild type (a) and various mutant pseudoknots (b-e) with their corresponding frameshifting efficiencies. Mutations as compared to the wild type are boxed. The two stabilizing residues at the 3'-end of pk103 are drawn as open characters. The three boxed residues in L2 highlighted in a and b have been deleted in c-e, the boxed base pairs in S1 and S2 as well as C20 (d,e) indicate mutations with respect to the wild type.*

frameshift efficiency close to the wild type level (20%). This suggests that the identity of the bases in the SRV-1 L2 are not important. Therefore, iii) we changed G20 into a cytidine, thus destabilising the putative pseudoknot-dimer interaction, and indeed for this mutant, pk102, no concentration-dependence behaviour of the imino proton resonances was observed. Frameshift efficiency for the triple mutant pk102 (Figure 5.1d) was again comparable to wild type level, 24%.

However, pk102 showed an increased intensity of two broad peaks in the region of the NMR spectrum where the non-hydrogen bonded imino protons tend to resonate (10.5-11.5 ppm), which indicated aggregation of a different kind. This multimerization effect could be attributed to the four consecutive guanosine residues at the 5'-side of the RNA pseudoknot. These base pairs are probably important for function: in almost all frameshifting pseudoknots, S1 starts with 3 or 4 G-C base pairs, and mutational studies showed that changing the third base pair decreased frameshift efficiency by approximately 50% (ten Dam *et al.*, 1994). To be able to study a monomeric RNA pseudoknot iv) we

changed the second base pair of S1 into a C-G base pair, thereby interrupting the stretch of four guanosines. This mutation resulted in a complete disappearance of the two broad imino-proton peaks around 11 ppm supporting our hypothesis that the multimerization was due to the four consecutive guanines.

Finally, evaluation of a few additional H5/H6 crosspeaks with minor intensities (approximately 10%) in the TOCSY experiment pointed out the presence of a species with a local alternative conformation, originating from heterogeneity at the 3'-end. The purified n+1 transcript showed less of the undesired minor conformation, which is probably caused by stacking effects of the additional single stranded nucleotide at the 3'-end. Therefore, two residues, an adenosine and a cytidine, were added to the 3'-end of the pseudoknot, thus favouring the major conformation. The final mutant, pk103, combined all necessary changes for studying the monomeric form of the SRV-1 pseudoknot by NMR, while maintaining an adequate level of frameshift efficiency (16%, Figure 5.1e).

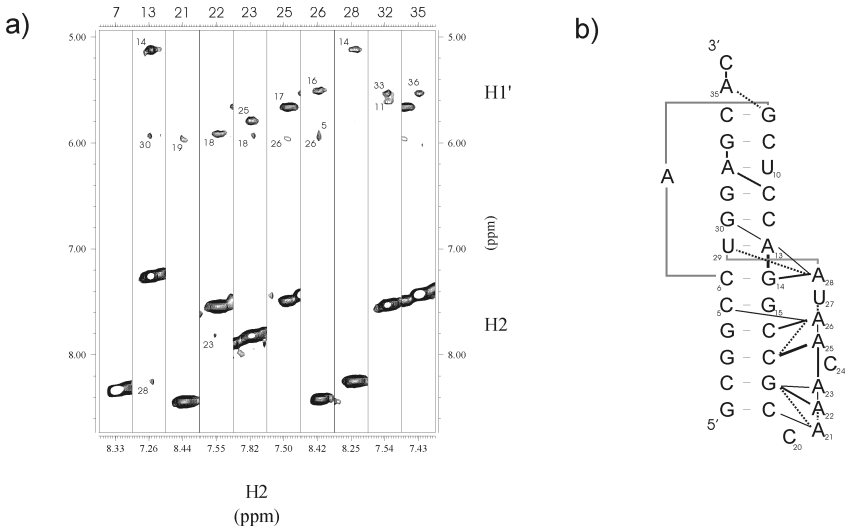

Figure 5.2 *a) Strips from a 3D NOESY-HMQC experiment belonging to all H2 protons of pk103 showing long range NOEs to other H2 and H1' protons. b) Overview of all NOEs from H2 to H1' protons observed in 2D NOESY and 3D ^{1}H-^{13}C NOESY-HMQC spectra. The thickness of the lines corresponds with the intensity of the NOEs. The dashed lines correspond to the weak NOEs only visible in the 2D NOESY.*

Resonance assignment and structure determination

The secondary structure of pk103 - which turned out to be identical to the published model (ten Dam *et al.*, 1994) - was evident from a complete sequential NOE walk, connecting all base paired imino protons with the exception of G1 (typical for a 5'-fraying residue) and the step between G14 and U29 at the S1-S2 junction. Despite the absence of a NOE between the slightly broadened imino proton resonances of residues G14 and U29 - most likely due to dynamic properties of the junction and enhanced chemical exchange of the imino protons - the assignment of the imino proton of U29 and the presence of a U-A base pair at the junction could be established unambiguously. While a single adenine spans the major groove, still a clear sequential NOE contact was observed between G8 H1 and G33 H1 at the 3'-end.

Standard assignment procedures (Wijmenga *et al.*, 1993, Wijmenga and Van Buuren, 1998), using homo- and heteronuclear experiments in D_2O led to a near complete assignment of the non-exchangeable protons. The sequential walk from anomeric to aromatic protons was interrupted between residues 6-8, 19-20, in the middle of L2 at residues 23-24 and at residues 28-29, indicative of a change in direction of the backbone or increased flexibility in these regions. Strong intra-residue NOEs between H1'/H2'/H3' and H6 of cytidine 24 suggested higher flexibility for this residue in comparison with the rest of the loop residues. This was confirmed by initial $T_1\rho$ relaxation experiments, which indicated an increased flexibility for this residue. All residues involved in Watson-Crick base pairs showed typical A-type helical NOE contacts. A structurally important "pseudo-sequential" NOE contact was observed between C6 H2' and U29 H5 and H6 which helped restraining the relative orientation of the two stems.

Sequential NOEs, involving ribose and aromatic C6/C8 protons observed for practically all residues in L2 indicated stacking in an A-type fashion. Assignment of the loop 2 adenine H2 resonances was obtained using a 3D HCCH-TOCSY (Legault *et al.*, 1994), which connected each H2 with the already assigned C8 protons. This allowed for resonance assignment using the large number of NOEs observed between the H2s and the H1's of different residues in S1 and L2. In Figure 5.2a strips of all H2 resonances in the 3D NOESY-HMQC are shown with NOEs to other H2 and H1's.

The high adenine content of L2, six out of nine residues, and the large number of NOEs turned out to be very important in the structure determination. All L2 adenine H2s

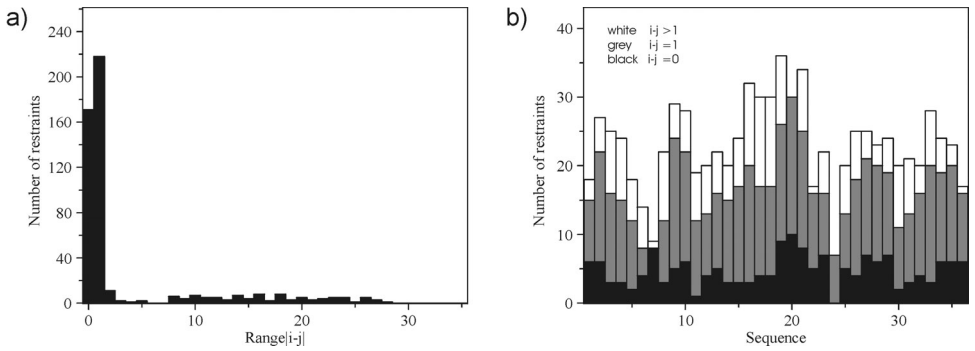

Figure 5.3 *a) Bar graph of all distance restraints used in the structure calculation. a) Representation of the number of distance restraint between residue i and j. b) Representation of the number of distance restraints and the range for the individual residues. Intraresidual distance restraints are colored black, sequential grey and everything beyond sequential is colored white. Bar graphs are generated using DYANA (Güntert et al., 1997) and the number of restraints, as counted by DYANA, are devided by a factor of two resulting in the "real" values as used in the structure calculation.*

showed sequential NOEs to H1's in L2, and a number of long range NOEs, predominantly to the S1 strand running anti-parallel to L2 (schematically displayed in Figure 5.2b). Figure 5.3 shows a schematic presentation of all 547 NOE derived distance restraints that were used together with 272 dihedral restraints. For the structure calculations the torsion angle dynamics approach implemented in the Xplor program (Brünger, 1992) was used as described by Stein *et al.* (1997). Mainly due to the large number of distance restraints between L2 and S1, the structure calculations resulted in a very compact and well defined pseudoknot structure as shown by the superimposed structures in Figure 5.4.

Description of the structure

A single adenine spans the major groove of S2

The S2 stem consists of 6 base pairs. Its major groove is spanned by the single residue, A7, with the base buried deeply into the groove (Figure 5.5a). The sugar pucker of A7 is of the S-type and the γ torsion angle is trans instead of the normal gauche$^+$, resulting

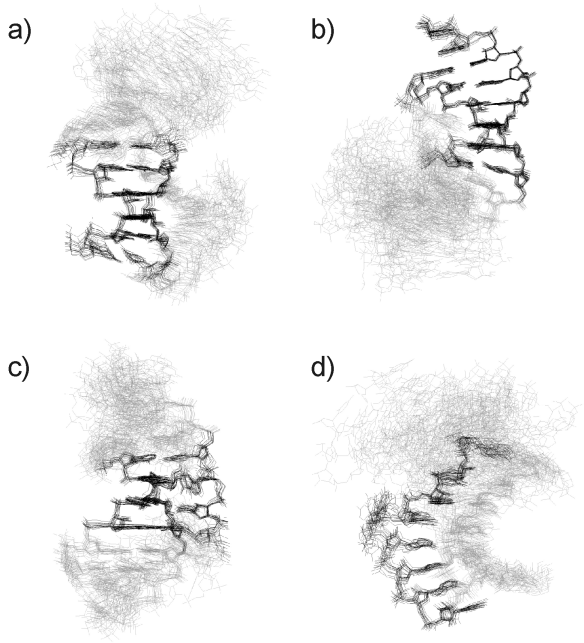

Figure 5.4 *Heavy atom superpositions of the 15 lowest energy structures fitted to: a) S1 residues 1-6/14-19 (rmsd = 0.5 Å); b) S2 residues 8-13/29-34 (rmsd = 0.3 Å); c) junction residues 13-16/4-6,29 26-28 (rmsd = 0.6 Å); d) L2 residues 20-28 (rmsd = 1.3 Å).*

in an average elongated phosphate-phosphate distance of 6.8 Å between residue 7 and 8. Similar a-typical torsion angles were observed in the TYMV pseudoknot for extending one of the two residues of L1, (Kolk *et al.*, 1998b). However, the phosphate-phosphate distance over 6 base pairs in an A-type helix is approximately 11Å (Pleij *et al.*, 1985, Haasnoot *et al.*, 1986), which is too large for a single residue to span. To compensate for this effect, the distance is shortened by a large helical twist of 49° between the two base pairs at the junction (vide infra).

Continuous stacking and base triple interactions form a highly structured L2-S1 interface

No less than 36 long range contacts were determined between the C2 protons of all adenines in L2 and the base pairs in S1. The large number of NOE derived distance

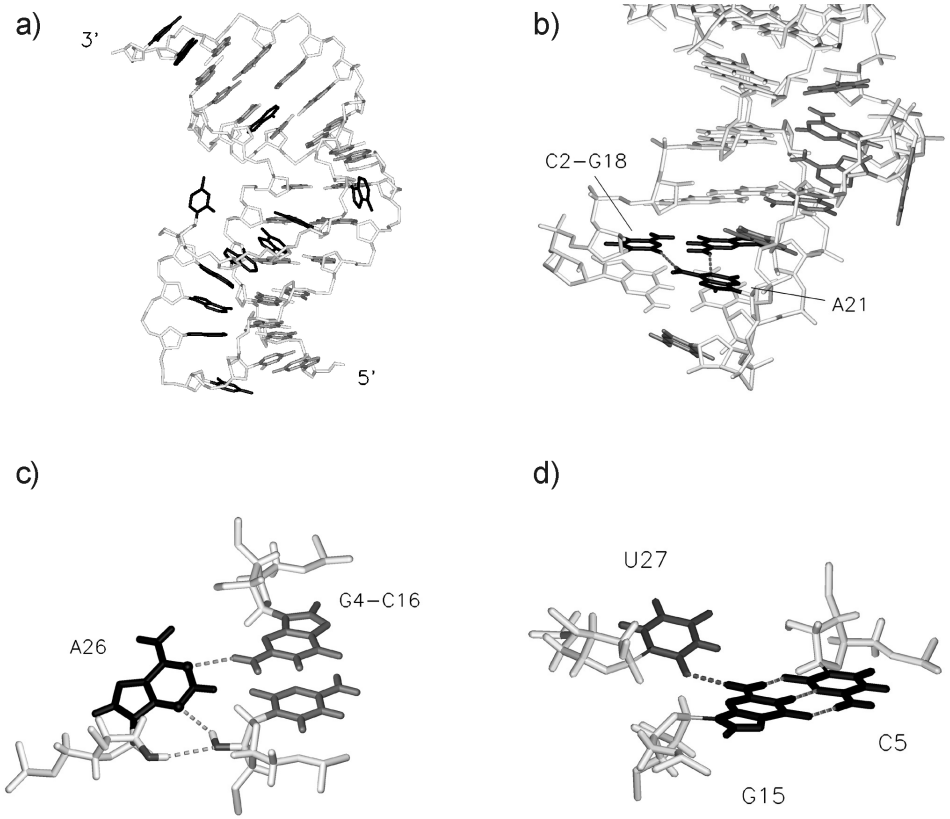

Figure 5.5 *a) Lowest energy structure of the SRV-1 derived pseudoknot causing ribosomal frameshifting. The bases of the two stems involved in Watson-Crick base pairs are colored grey, the adenine comprising loop L1, the bases of loop L2 and the two 3'-stacking bases are colored black. b) Detailed view of the 5'-side of the SRV-1 pseudoknot with the bases C20-U27 of loop L2 colored in grey. Adenine 21 has two hydrogen bonds to the second base pair in stem S1, indicated by the dashed lines, and serves as a platform for consecutive stacking of the following adenine residues. c) Detailed view of the triple interaction between A26 of loop L2 and the G4-C16 base pair from stem S1. Three hydrogen bonds are present, indicated by the dashed lines. A26 is directed into the minor groove and forms a hydrogen bond by its N1 and the amino proton of G4. A ribose zipper motif was observed between the other two hydrogen bonds from the N3 and the hydroxyl group of A26 to the hydroxyl group of C16. d). Triple base pair formed by U27.C5-G15. Uridine 27 is held in the minor groove by a hydrogen bond between O2 and the amino proton of guanine 15.*

restraints results in a very well defined structure of L2, which is locked in the minor groove of S1 by various interactions. At the 5'-end of L2, at C20, the backbone turns sharply into the opposite direction and is held in position by A21, which forms two hydrogen bonds between its N1 and the amino proton of G18, and between its amino proton to the O2 of C2 (Figure 5.5b). Subsequently, A21 serves as a platform for consecutive stacking of all residues up to U27, with the exception of C24. A23 and A25 are stacked - evidenced by NOEs from the H2 of A23 to the H2 and H1' of A25 - , while C24 is bulged. The dynamic properties of C24 (visualized in Figure 5.4d) are transposed to the next residue and causes line broadening of the aromatic proton resonances of A25. However, in the calculated structures A25 is firmly held in position close to the minor groove connected by a strong NOE between its H2 and C17H1' (Figure 5.2a). No hydrogen bonds can be detected between residues 22-25 and S1, so it appears that the structure of this part of L2 is primarily formed by stacking interactions between the adenine bases.

At the 3'-side of the loop two base triples are present. Adenine 26 is stacked upon residue 25 and forms a base triple with G4-C16 in S1 (Figure 5.5c). Additionally, a single ribose zipper motif (Cate *et al.*, 1996) is present between A26 and C16, where the 2'-hydroxyl group and N3 of A26 share a hydrogen bond with the 2'-hydroxyl group of residue C16. The other base triple is formed between U27 and G15-C5, stabilized by a hydrogen bond between U27 O2 and the amino group of G15 (Figure 5.5d).

The large helical twist between S1 and S2 at the junction together with an S-turn at the 3'-side of L2 allow the three strands of S1 and L2 to come in close proximity at the junction and positions the 3'-side of L2 in the minor groove for forming the base triples (Figure 5.6a). A28 fulfils a key role in maintaining the structural arrangement at the junction. The S-turn is held in place by the rather unusual position of the A28 base (Figure 5.6b), which is fixed perpendicular to the plane of the canonical base pairs in the stem region, stabilized by hydrogen bonds between the N1 and the amino protons of A28 and the 2'-hydroxyl group of G14 (Figure 5.6a). The ribose moiety of A28 is tilted by about 90° bringing the H1' of A28 and U29 in close proximity (Figure 5.6a). Figure 5.6a also clearly shows that the sharp turn between A28 and U29 and the tilted ribose of A28 positions U29 in such a way that it can base pair with A13 in a canonical manner without disturbing the base triple below. Finally, the phosphate group between A28 and U29 is in close proximity

to the 2'-hydroxyl group of C6 in the opposite strand, which could possibly form a hydrogen bond to stabilize the unusual conformation of the backbone around A28.

Discussion

Structural comparison

This work presents, after elucidation of the TYMV (Kolk *et al.*, 1998a) and BWYV (Su *et al.*, 1999) pseudoknots, the third structure of an H-type or classical pseudoknot (for a recent review discussing nomenclature of pseudoknots see Hilbers *et al.*, 1998) solved at sufficient resolution to provide detailed structural information. The structure of the frameshifting RNA pseudoknot of SRV-1 exhibits the typical topology seen for other previously studied H-type pseudoknots structures. At the local structural level, however, there are a number of interesting differences and similarities, which we will elaborate below.

The present SRV-1 pseudoknot structure exhibits an unexpectedly compact conformation with several tertiary interactions, mainly at the S1-L2 interface (Figure 5.4)

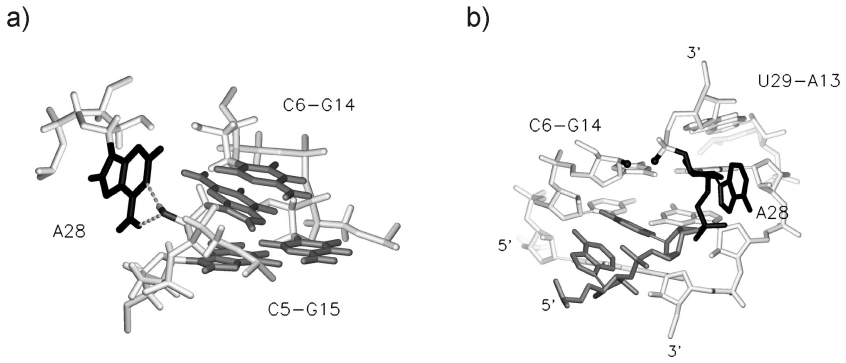

Figure 5.6 *a) Adenine 28 (colored black) is oriented perpendicular to the plane of base pairs (colored grey) and is hydrogen bonded by N1 and the amino proton to the hydroxyl group of guanine 14. b) Representation of the heavy atoms of the junctional region (G4-C6, A13-C16, A26-U29). The tilted ribose of adenine 28 (colored black) changes the direction of the backbone in such way that uridine 29 can be base paired with adenine 13. The two oxygen atoms (small black spheres) might be involved in a water mediated or direct hydrogen bond.*

similar to what has been observed for the TYMV and BWYV pseudoknots. Most of the interactions between S1 and L2 consist of base triple formation between adenines in the loop and the bases of the base pairs in the stem. We want to mention the one exception namely the A26:G4-C16 base triple, which is stabilized by a single base zipper motif (Figure 5.5c). Intramolecular ribose zippers, unique to RNA since hydroxyl groups are involved, have previously been observed in the P4-P6 domain of Group I intron ribozymes (Cate *et al.*, 1996), the HDV ribozyme (Ferré-D'Amaré *et al.*, 1998) and the L11-binding domain of 23S rRNA (Conn *et al.*, 1999, Wimberly *et al.*, 1999). In these structures at least two pairs of riboses interact by hydrogen bonding to form the ribose zippers, which are characterized by shared hydrogen bonds between the 2'-hydroxyl group and a pyrimidine O2 (or purine N3) of one nucleotide and the 2'-hydroxyl of its partner. In the SRV-1 pseudoknot the riboses of the next residues in the stack, U27 and G15, are not participating in the ribose zipper. Therefore, in a more general sense the hydrogen-bonding pattern found here for a single ribose pair might be considered the signature of a ribose zipper motif.

A pertinent finding in this work is that the base pairs predicted in the secondary structure model of the SRV-1 pseudoknot (ten Dam et al., 1990) are formed in the three-dimensional structure. Different views, however, have been articulated for the presence of the U29-A13 base pair at the junction. In the Mouse Mammary Tumor Virus (MMTV) pseudoknot an unpaired adenine was found at the junction, leading to a bent pseudoknot, which has been proposed to be an essential structural feature for frameshifting (Shen and Tinoco 1995, Chen *et al.*, 1996). The MMTV pseudoknot and variants thereof with an intercalated adenine base at the junction showed frameshift efficiencies of about 20%. For variants without an intercalated adenine the frameshift efficiencies were much lower (2%). Functional data (Chen *et al.*, 1995) led to the proposal that the SRV-1 pseudoknot would, at the dispense of the U29-A13 base pair, also contain an unpaired intercalating adenine at the junction to create this characteristic bent conformation (Chen *et al.*, 1996). Mutational studies (Sung and Kang, 1998) of the SRV-1 pseudoknot in which the uridine had been changed into all other three nucleotides without affecting frameshift efficiency further supported this hypothesis. In our work, however, we firmly assigned a hydrogen bonded U29 imino proton and unambiguously established the presence of a closed U29-A13 base pair at the junction. Earlier work by Hoffman and coworkers - who were able to assign the U29 imino proton - also indicated the presence of this base pair (Du *et al.*, 1997). Thus, an

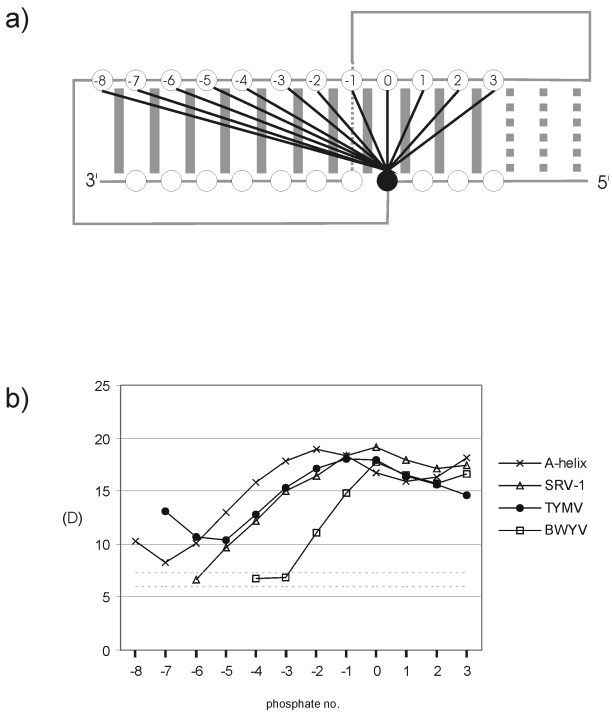

Figure 5.7 *Interstrand phosphorus-phosphorus distances in A-form RNA (Schindelin et al., 1995) and pseudoknots. The curves (b) represent the distances between the sampling phosphate in one strand (indicated by the black dot in (a)) and the phosphates (numbered from –8 to 3 in (a) and (b)) on the other strand. For the A-form helix the choice of the sampling phosphate is arbitrary, for the pseudoknots the sampling phosphate is 5' to the first L1 nucleoside. The dashed lines in (b) indicate 6.0 and 7.4 Å, the sequential phosphate-phosphate distance in a regular A-form helix and the maximal distance that can be spanned by one nucleotide, respectively.*

unpaired, intercalated nucleotide at the junction is not necessarily required for efficient frameshifting of the SRV-1 pseudoknot.

Geometrical aspects of the classical pseudoknot architecture

In the H-type pseudoknots only a small number of nucleotides, i.e. one or two, span the major groove of S2. The number of base pairs that can be bridged in this way has been a matter of debate. For instance, NMR studies of the pseudoknot in gene 32 mRNA of

bacteriophage T2 showed that a single nucleotide can span the major groove of a six or seven base pair stem, but not five or eight (Du and Hoffmann, 1997). On the other hand, it has been observed that in the BWYV pseudoknot only one nucleotide spans the major groove of S2, which effectively consists of four base pairs. Using the SRV-1, TYMV and BWYV structures we now further elaborate on these considerations.

To this end the interstrand phosphorous-phosphorous distances from the 3'-end of S1 - defined as the sampling phosphate (Figure 5.7a) – to the opposite strand in S2 are compared with the distances expected for a continuous A-type RNA double helix (Figure 5.7b). This graph shows that for canonical A-type RNA the shortest phosphorous-phosphorous distance (~8Å) occurs after a build-up with 7 base pairs, thus leaving a gap that can be bridged by one (or two) nucleotides. The distances measured in the TYMV and SRV-1 pseudoknots nearly coincide. In the SRV-1 pseudoknot a distance of 6.8Å is reached when S2 is six base pairs long. This distance can easily be bridged by one nucleotide. A similar situation obtains in the TYMV pseudoknot. Here S2 is five base pairs long and the remaining gap, 11Å, can be spanned by two nucleotides, the effective length of L1. In the BWYV pseudoknot the situation is significantly different. Here the phosphorous atoms of the third and fourth base pair of S2 are already at a distance of 6.5Å from the sampling phosphate. Again this distance can be, and in fact is spanned by one nucleotide. Below it will be discussed that the twist angle between the two base pairs at the pseudoknot junction is an essential element in imposing these differences.

The helical twist between the base pairs in an idealized A-type RNA helix is 32.7º. Compared to this value large helical twists at the junction between S1 and S2 seem typical for the formation of classic pseudoknots. Thus, for the SRV-1 pseudoknot a helical twist of 49º is observed between the two base pairs at the junction. In the TYMV-pseudoknot this twist amounts to 52° (Hilbers *et al.*, 1998) and for the BWYV pseudoknot (Su *et al.*, 1999) a twist of about 90° has been deduced (Hilbers *et al.*, 1998). The large helical twist in these pseudoknots serves two purposes. First, it facilitates bridging of the major groove of S2 by a small number (one or two) of nucleotides. Secondly, it opens the minor groove at the junction to accommodate base triple formation in the TYMV (Kolk *et al.*, 1998a, Hilbers *et al.*, 1998) and the SRV-1 pseudoknots (Figure 5.4) or base quadruple formation in the BWYV pseudoknot (Su *et al.*, 1999). Formation of the base quadruple in the latter molecule also accounts for the large difference with the proposed secondary structure model.

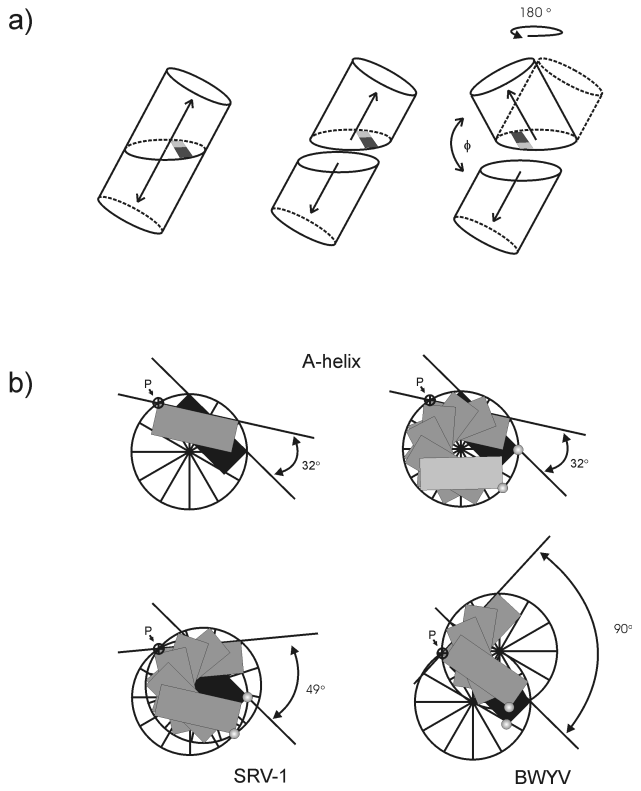

Figure 5.8 *a) Schematic representation of creating a helical bend (ϕ) by overwinding the helical twist of one base pair step to 180º. b) Schematics of base stacking in helical A-form RNA and SRV-1 and BWYV pseudoknots. Base pairs are represented by rectangles with the C1' of standard Watson-Crick base pairs at the corners touching the sugar-phosphate backbone, indicated by the circle. Helical twists are indicated as the angle between adjacent C1'-C1' vectors across base pairs.*

Using a very simple model it can be demonstrated that overwinding the helical twist of one of the base pairs can induce a helical bend. This is shown in Figure 5.8a, where a cylinder serving as a model for an RNA helix is intersected by a plane at a certain angle representing the tilt of the base pair. Subsequently, the upper part of the helix (cylinder) is rotated by 180º around an axis perpendicular to the plane through the mentioned base pair, taking the point where the helix axis crosses the plane as the pivot point. If one takes the

base pair tilt equal to 45º it can be envisaged easily that the combined large twist and tilt angles lead to a bend of 90º between the upper and lower part of the helix. It is also clear from this example that smaller tilts will result in smaller bends.

At this point we like to mention that, interestingly, superposition of the pseudoknot structures on that of a regular A-type helix shows that in the pseudoknots the deviations of the local as well as the global helix axes from those in the regular A-type helix are not as outspoken as suggested in earlier publications. For this one has to realize that the number of base pairs in the helical stems, S1 and S2, in the pseudoknots do not entail a complete helical turn suggesting that the reported bends refer to local instead of global bends. To appreciate the importance of this remark a comparison with the situation in an A-type helix is enlightening. If one takes a perfect A-form helix of twelve base pairs the apparent bend between the helix axis of the upper and lower 6 bp stems, generated by either program Curves (Lavery and Sklenár, 1988, 1989) or NEWHEL93 (R.E. Dickerson, University of California, Los Angeles), amounts to ~22º. Thus, examination of two stacked helical A-type stems smaller than one helical turn by either these programmes or by eye inspection may lead to the conclusion that the molecule is bent while actually the global helix is straight. This does not mean that in pseudoknot structures no bending of the global helix axis occurs at all. It is, however, in our opinion rather modest.

The overwinding with the pivot point located at the centre of the helix is, however, not possible in practice. It would require a lengthening of the phosphorous-phosphorous distance, which is prohibited. In fact, variation of the distance between two consecutive phosphorous atoms in the same strand is only possible within small margins. In nucleic acid structures this is reflected in the presence of almost constant phosphorous-phosphorous distances (Hilbers *et al.*, 1994). This means that in reality the rotation axis has to be moved close to the cylinder wall. Then rotation of the upper part of the helix will leave the orientation of its helix axis unaltered with respect to the same rotation in the situation above. However, this rotation will now result in a shift of the helical centre of the upper with respect to that of the lower helix in the intersecting plane. This is illustrated in Figure 5.8b, where the effect of overwinding to a helical twist of 49º and 90º in the SRV-1 and BWYV pseudoknot, respectively, is compared with the situation in a normal A-type helix. In Figure 5.8b the black rectangle represents the last base pair of the lower helix, and the grey shaded rectangles represent base pairs in the upper helix stacked upon the last base

pair of the lower helix. Build up of the upper helix in the normal A-type helix by seven base pairs brings, in this two-dimensional projection, the 5'-terminus of the upper stem – indicated by the little sphere attached to the upper base pair - close to the 3'-terminus of the lower stem – indicated by the little sphere attached to the black base pair. This results in a distance of ~ 8 Å between these terminal phosphates (c.f. Figure 5.8b). In the SRV-1 pseudoknot the twist angle between the upper base pair of the lower helix (stem S1) and the lowest base pair in the upper helix (stem S2) amounts to 49º giving rise to a shift of the helix centre of the upper helix, as can be seen from shift of the upper helix with respect of that of the lower helix. Note that the rotation is around the pivot point P located at the crossing of the two circumferences, leaving the distance of the consecutive phosphorous atoms of the upper base pair of stem S1 and the lowest base pair of stem S2 virtually unaltered. A build up of the upper helix by 6 base pairs brings its 5'-terminus as close to the 3'-terminus of the lower stem as the similar build up by 7 base pairs in the A-type helix. This clearly underscores the proposition that increasing the twist angle shortens the distance between the two termini in nice correspondence with the results depicted in Figure 5.7b. The effect is even more dramatic in the BWYV pseudoknot. A twist angle of 90º between the base pairs at the pseudoknot junction leads to a much larger shift between the centres of the helices, leading to a short distance of the free termini of the upper and lower stems after a build up with 3 base pairs and the stacking adenine. Again this nicely conforms to the experimental results in Figure 5.7b.

Comparison with frameshift efficiencies

The structural features of the SRV-1 pseudoknot nicely fit the results of mutational studies on frameshift efficiencies and provide leads for further investigations. The absence of interactions between the single L1 residue A7 and the major groove of S2 strongly suggests that neither the identity of the crossing nucleotide nor the base pair composition of S2 is important for frame shifting. This fully agrees with the various SRV-1 pseudoknots mutated in this region, that only show slight changes in frameshift efficiencies (Figure 5.9, ten Dam *et al.*, 1995). The only known exception, i.e. a drop of 13% upon changing two G-C pairs into A-U pairs supports the notion that base pair composition of S2 is not important provided that stability is maintained. The intricate interactions at the S1-S2 junction and at the L2-S1 interface suggest that in this part of the pseudoknot the nucleotide composition is

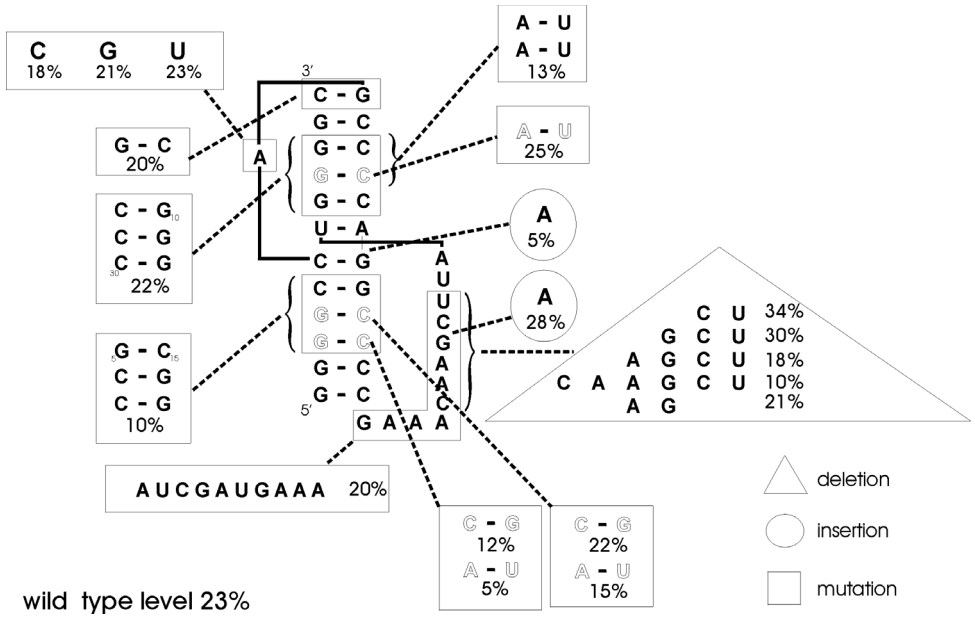

Figure 5.9 *Frameshift efficiencies of selected SRV-1 pseudoknot mutants (ten Dam et al., 1996).*

much more important. Again this concurs with the various mutational studies performed. The wild type sequence at the 3'-side of L2, CUUA, is apparently suboptimal in terms of efficiency, since deletion of 2 (CU) or 3 (GCU) residues from L2 was shown to increase frameshift efficiency to a level of 34% and 30% respectively (ten Dam *et al.*, 1995). This suggests that the interactions at the L2-S1 interface, i.e. base stacking, base-triples and the ribose zipper motif, are probably important elements in the mechanism of ribosomal frameshifting. In other words, absence of the well-defined structure in the wild-type pseudoknot at the junction might explain the decrease in frameshift efficiency. When AG, which does not show direct contacts with S1, was deleted from the loop, frameshift efficiency was slightly less (21%) than the wild type level (23%). In the case of deletion of CU, a guanosine is left between A26 and U27. Here the higher frameshift efficiency might be due to additional stacking interactions of the guanosine between A26 and U27 and a triple interaction with one of the base pairs in S1. Further shortening of L2 progressively diminishes frameshift efficiency, which can be the result of disruption of the prime

interactions at the junction by tension of a too short L2. When a block of three base pairs in S1 is inverted, the frameshifting efficiency drops to 10% (ten Dam *et al.*, 1994), which can be attributed to the loss of two triple interactions. For our NMR studies we introduced a single base pair inversion at G2-C18. In the NMR structure the sharp turn of L2 at the base of S1 is held by a hydrogen bond between the N1 and the amino proton of A21 and the amino proton of G18 and the O2 of C2, respectively. Although it can be envisaged that in the wild-type pseudoknot a similar hydrogen bond can be formed by shifting A21 slightly over in the L2 minor groove, this minor change might decrease the overall frameshift efficiency to the respectable value of 16%.

In summary, nearly all the results of the mutational studies can be explained by our structure. However, the single base pair inversion of G3-C17 decreases the frameshift efficiency to 12% (ten Dam *et al.*, 1995), while this base pair is not involved in any interaction in our structure. This clearly shows the need for further structural and functional studies to fully comprehend the importance of the structural details of the SRV-1 pseudoknot. In particular, it will be very important to see whether the tertiary interactions at the junction as found in the pseudoknot structure of the SRV-1 derivative presented here are also present in the wild type pseudoknot. Also, to date the base triples and other interactions at the junction have not been put to a functional test yet.

The importance of adenines

It is clear from the structures of the SRV-1 pseudoknot, presented here, and the BWYV pseudoknot (Su *et al.*, 1999), that adenines can be ascribed a special role in ribosomal frameshifting. Insertion of an adenine between G26.1 and C26.2 in L2 increases frameshift efficiency to 28%, a result for which no explanation could be given before (ten Dam *et al.*, 1995). Following earlier arguments, this particular adenine might also be involved in stacking and base triple interactions, giving rise to the remarkable high frameshift frequency. Alternatively, the changes in frameshift efficiency upon deletion or insertion of nucleotides in L2 could possibly arise from an optimization of the loop length as well as of the adenine content. A similar result has been observed in BWYV, where the insertion of an adenine residue after the first residue in L2 (G19A19a) increased frameshift efficiency by 70% (Kim *et al.*, 1999).

It would be interesting to investigate the effect on the L2 backbone conformation and frameshift efficiency following replacement of adenine 28 by a pyrimidine. A purine at position 28 (Figure 5.6) seems important, since the single aromatic ring of a pyrimidine will probably be too far away to form a hydrogen bond with the 2'-hydroxyl group of G14. In line with this observation Brierly and coworkers (Liphardt *et al.*, 1999) found for an infectious bronchitis virus (IBV) derived pseudoknot an adenine at the 3'-end of an 8nt L2 essential for efficient ribosomal frameshifting. By increasing the size of the loop to 14nt, the specific requirement of this adenine is lost. However, the IBV pseudoknot is structurally similar to the MMTV frameshifting pseudoknot with 2 nucleotides in L1 and an intercalating adenine at the junction. Since our structure of the SRV-1 pseudoknot is distinctly different from the MMTV pseudoknot (Shen and Tinoco, 1995) it is not evident that when the three bases (GCU), deleted from loop L2 in SRV-1, are reinserted, the typical structural features and requirement of adenine 28 is lost in SRV-1 as well.

Various mutational studies conducted on MMTV (Chamorro *et al.*, 1992, Chen *et al.*, 1995) and IBV (Liphardt *et al.*, 1999) pseudoknots, compared with the SRV-1 mutation studies (ten Dam *et al.* 1994, 1995) have led to the impression that the structure of the SRV-1 pseudoknot would differ from the minimal IBV derived pseudoknot, TYMV or BWYV structures (Liphardt *et al.*, 1999, Giedroc *et al.*, 2000). In particular the absence of an effect of the available L2 mutations on SRV-1 frameshift efficiency led to the suggestion that the SRV-1 pseudoknot would not contain L2-S1 interactions as shown for the TYMV and BWYV pseudoknots (Liphardt *et al.*, 1999, Giedroc *et al.*, 2000). However, at the time of these studies frameshift data of critical mutants involving tertiary interactions at the stable junction were not available. Based upon our structural study we propose that the IBV derived pkA-A and minimal IBV pseudoknot (Liphardt *et al.*, 1999) have structures similar to the SRV-1 pseudoknot with similar L2-S1 interactions.

Implications for the ribosomal frameshift mechanism

The solution structure of the SRV-1 pseudoknot involved in ribosomal frameshifting presented here, has several special structural properties that seem very important for its function. Frameshifting appears to be a delicate mechanism that can be fine-tuned for any particular species in order to obtain the proper ratio between functional proteins needed in the life cycle of the virus. Base pairing as obtained by S2, stacking of the bases in L2 and

tertiary interactions between S1 and L2, seem to be used as a tuning device in order to regulate the ratio between the translational products of the overlapping reading frames. Beyond the A-type double helix, our poor knowledge on the structural properties of loops and bulges leave us unable in explaining the effect certain mutations have on the frameshift efficiency. Therefore it is important to understand the structural effect of subtle changes in the loops, such as the change of a single adenine in the IBV-based pseudoknot (Liphardt *et al.*, 1999). Further structural analysis by either NMR or crystallography not only of the wild type sequences, but also of mutant pseudoknots are called for.

Despite the large amount of structural and functional data the actual role of the pseudoknot structure in the mechanism of ribosomal frameshifting is still unclear (for a recent review see also Giedroc *et al*., 2000). Most confusing is the large number of pseudoknots differing in size, loop sizes and nucleotide composition, not to mention the hairpin structures and three way junctions that have been found to induce ribosomal frameshifting as well. Therefore, the discovery of a common feature present in all types of pseudoknots might provide a basis for a better understanding. In fact the only general feature that has been recognized so far for the frameshifting pseudoknots is the definition of formation of a classic pseudoknot itself: a single stranded region which folds back and forms base pairs with residues in a hairpin loop. Therefore, as a consequence of the pseudoknot structure, the translating ribosome proceeding along the mRNA will approach a hairpin structure and a third strand connected to the loop upstream. Unwinding of the pseudoknot might be difficult because of the stability of S2 and the loop-helix interactions, thereby leaving time for the ribosome to pause and shift to the –1 frame. Alternatively the specific interactions between L2 and the minor groove of S1 as seen in the SRV-1 pseudoknot that lead to the formation of a triple helix raises the intriguing possibility that a triple helix connected by a loop might be the recognition signal for a translating ribosome to frameshift.

Materials and methods

In vitro transcription and translation assay

Mutations made in the SRV-1 *gag-pro* frameshift signal discussed in this report were made starting from the pSF43 mutant previously described by ten Dam and coworkers

(1995). Plasmids were purified with the Wizard plasmid DNA purification system and used in the Promega TNT SP6 quick coupled transcription/translation system according to the manufacturer's protocol. The products of translation of the "in frame" and –1 reading frame, are a 19kD and 22kD fragment in size and contain 10 and 11 methionines respectively. Synthesised proteins were separated on a 15% SDS polyacrylamide gel. The relative amounts of the proteins from both reading frames were quantified by determining the incorporation of ^{35}S-methionine by scanning the gel with a Biorad phosphor-imager correcting for the background and the differential methionine content of the products.

Sample preparation

An unlabelled and four different residue-type 36-mer RNA molecules specific labelled in ^{13}C/^{15}N were prepared enzymatically by in vitro transcription using T7 RNA polymerase (Milligan *et al.*, 1987) with a DNA template. The products were purified using preparative polyacrylamide gel electrophoresis and recovered by electroelution. Purified RNA was extensively dialyzed against 100 mM NaCl, 10 mM $KHPO_4/KH_2PO_4$ buffer, pH 6.4 and concentrated using Centricon microconcentrators. The final concentration of the unlabelled sample was 3 mM and 1-2 mM for the residue specific ^{13}C/^{15}N labelled RNA samples. Addition of 6 mM magnesium did not alter the chemical shifts or NOE patterns, indicating that the presence of magnesium does not influence the pseudoknot structure.

NMR methods

All spectra were recorded on Varian Inova 500 MHz and 750 MHz and Bruker DMX 600 MHz spectrometers, at 10° C for samples in H_2O and at 30°C for those in D_2O. Exchangeable protons were assigned from 2D NOESY (Jeener *et al.*, 1979) experiments which were recorded using a jump return pulse sequence (Plateau and Guéron, 1982) for water suppression or watergate pulse sequence with a water flip-back pulse (Lippens *et al.*, 1995). Non-exchangeable protons were assigned using 2D NOESY spectra employing mixing times of 50, 100, 200 and 300 ms, and from DQF-COSY (Shaka and Freeman, 1983) and TOCSY (Griesinger *et al.*, 1988) experiments. Phosphorus chemical shifts were assigned by a 2D ^{1}H-^{31}P Hetero-TOCSY-NOESY (Kellog and Schweizer, 1993) and a ^{1}H-^{31}P HETCOR experiment (Sklenár *et al.*, 1986).

Four samples with the different residue types separately $^{13}C/^{15}N$ labelled were used to separate overlapping resonances and ^{1}H-^{13}C HMQC (Bax *et al.* 1983), ^{1}H- ^{13}C 3D NOESY-HMQC (Ikura *et al.*, 1990) and HCCH-TOCSY spectra (Bax *et al.*, 1990, Legault *et al.*, 1994) were recorded. Assignments of the ^{1}H, ^{13}C and ^{31}P resonances are listed in the Appendices chapter in tables A.*IV* and A.*V*.

Structural restraints

Proton distances involving non-exchangeable protons were predominantly estimated from 100, 200, and 300 ms 2D NOESYs as well as a 300 ms 3D NOESY-HMQC. NOE crosspeak volumes involving non-exchangeable protons from the 100 ms NOESY were converted into distance restraints using the approach as described by Barsukov and Lian (1993), which qualitatively takes care of spin diffusion, by using the H5-H6 NOEs (2.45 Å) for short and intraresidue H1'-H6/H8 NOEs of helical residues for longer distances (3.65 Å on average) for those crosspeaks that were only visible at longer mixing times. By this method, the distances that are derived from the 100 ms spectrum agree very well with those from longer mixing times, and the possibility of large errors exceeding the error bandwidth can be virtually excluded. Initially we made use of ±20% error bounds for distances derived from the 100ms NOESY experiment, because in our experience this gives a higher convergence rate, while the number of violations is kept to a minimum. Additional tests, using larger (-30%/+40%) error bounds resulted in structures with virtually identical overall and local folds, albeit with slightly larger overall rmsd (1.8 Å) and number of violations. Small NOEs taken from the 300 ms 3D NOESY-HMQC were subdivided in four classes strong (1.8-3.0 Å), medium (2.0-4.0 Å), and weak (3.0-5.0 Å). Small NOEs from the 200 ms NOESY were classified as very weak (4.0-6.0 Å). As spin diffusion might become important for a molecule of this size, the upper bound for very weak NOEs was set to 6 Å, which is a rather conservative estimate. Distance bounds for partially overlapping peaks were set to 1.8-6.0 Å. Non-NOEs (i.e. lower bounds of 4.5-5 Å between a given proton pair) were put in at later stages of the structure calculations when the distance between a given proton pair came out smaller than 5 Å, while there was clearly no crosspeak visible in any of the NOE spectra. The given lower bound is very conservative, and had no influence on the calculated structures, but merely helps defining the allowed conformational space during the molecular dynamics runs. Non-NOEs were only included

for residues for which we established separately that they are not involved in intermediate exchange. Scalar couplings between H1' and H2' were determined from a DQF-COSY and a NOESY experiment, and were converted to N- or S-type sugar puckers (using dihedral restraints) in case of small (<3 Hz) or high (>7 Hz) values, respectively. The sugar puckers of A21, A22, A23, C24 and C36 were left unrestrained because intermediate scalar couplings of 4-6 Hz were measured for these residues. All residues had intra residual H1'/H2'/H3' to H6/H8 NOE intensities that are indicative of an χ-angle in the *anti* range, except for C24 and were therefore restrained 202±30° for the stem residues and 202±60° for the others.

Both stem regions adopt an A-helical conformation, evidenced by chemical shifts, sugar puckers and NOE connectivities. All residues within the stem regions were therefore assigned glycosidic and backbone torsion angles in accordance with canonical A-type RNA with error bounds of ±20° for those residues. Distances relating to imino protons were conservatively estimated from a 300 ms NOESY spectrum. Hydrogen bond restraints - according to Saenger (1984) with error bounds of 0.1 Å - and planarity restraints were imposed for residues involved in Watson-Crick base pairing, as evidenced by the imino proton spectra and all observed NOEs and chemical shifts. Hydrogen bond restraints for the two junctional base pairs were loosened (i.e. error bounds were increased to 0.5 Å) in order obtain a more unbiased result in this region of the RNA molecule. No standard restraints were imposed for any of the backbone torsion angles between the sugar of C6 to G8, and that of C19 to U29, to ensure an unbiased course of the polynucleotide chain. For these domains, the β and γ torsion angles were left unrestrained, except for A29, for which we found evidence in the ^{1}H-^{31}P HETCOR experiment (Sklenár *et al.*, 1986) to restrain β to 0-120º. The angle ε was set to 225±60°, in order to exclude the stereochemically forbidden *gauche*$^{+}$ region. No downfield shifted phosphorous chemical shifts were found for any of the pseudoknot residues and therefore the angles α and ζ were conservatively restrained to 0±120° to exclude the trans region.

Structure calculations

A set of 100 structures was calculated using the torsion angle dynamics (TAD) protocol (Stein *et al.*, 1997) in X-PLOR (Brünger *et al.*, 1992). The TAD cooling step in this protocol was increased to 90 ps so as to obtain a higher convergence rate. 15 % of the

obtained pool of structures had no distance and dihedral restraint violations larger than 0.5 Å and 6°, respectively. All structures of this ensemble, best fitting the experimental data, were selected for presentation and their statistics are listed in the Appendices of this thesis Table A.*VI*. Color figures were generated using MOLMOL (Koradi *et al.*, 1996).

Protein Data Bank accession numbers

Coordinates for the set of 15 final structures as well as a full list of restraints used in Xplor have been deposited in the protein data bank, accession code 1E95.

References

References

Aboul-ela, F., Murchie, A.I., Norman, D.G. and Lilley, D.M. (1994), Solution structure of a parallel-stranded tetraplex formed by d(TG4T) in the presence of sodium ions by nuclear magnetic resonance spectroscopy, *J. Mol. Biol.*, **243**, 458-471.

Abramovitz, D.L. and Pyle, A.M. (1997) Remarkable Morphological Variability of a Common RNA Folding Motif. The GNRA Tetraloop-Receptor Interaction, *J. Mol. Biol.*, **266**, 493-506.

Allain, F.H.-T. and Varani, G. (1995), Divalent metal ion binding to a conserved wobble pair defining the upstream site of cleavage of Group I self-splicing introns *, Nucl. Acids. Res.*, **23**, 341-350.

Ban, N., Nissen, P., Hansen, J., Moore, P.B. and Steitz, T.A. (2000), The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution, *Science*, **289**, 905-919.

Bartels, Ch., Xia, T.-H., Billeter, M., Güntert, P. and Wüthrich, K. (1995) The program XEASY for computer-supported NMR spectral analysis of biological macromolecules, *J. Biomol. NMR*, **5**, 1-10.

Barsukov, I.L., and Lian, L.Y. (1993) NMR of nucleic acids; from spectrum to structure. In Roberts, G.C.K. (eds), *NMR of macromolecules. A practical approach*. Oxford University Press, New York, pp. 315-357.

Bassi, G.S., Mollegaard, N.E., Murchie, A.I., von Kitzing, E. and Lilley, D.M. (1995), Ionic interactions and the global conformations of the hammerhead ribozyme, *Nature Struct. Biol.*, **2**, 45-55.

Battiste, J.L., Mao H., Rao, N.S., Tan, R., Muhandiram, D.R., Kay, L.E., Frankel, A.D. and Williamson, J.R. (1996), Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex, *Science*, **273**, 1547-1551.

Bax, A., Griffith, R. & Hawkins, B.L. (1983) Correlation of proton and nitrogen-15 chemical shifts by multiple quantum NMR, *J. Magn. Reson*., **55**, 301-315.

Bax, A., Clore, G.M., Driscoll, P.C., Gronenborn, A.M., Ikura, M. and Kay, L. (1990a) Resonance assignment of methionine methyl groups and chi 3 angular information from long range proton-carbon and carbon-carbon J-correlation in a calmodulin-peptide complex , *J. Biomol. NMR*, **4**, 787-797.

Bax, A., Clore, G.M. & Gronenborn, A.M. (1990b) ^{1}H-^{1}H correlation via isotropic mixing of ^{13}C magnetization, a new three-dimensional approach for assigning ^{1}H and ^{13}C spectra of ^{13}C-enriched proteins, *J. Magn. Reson.*, **88**, 425-431.

Beattie, T.L., Olive, J.E. and Collins, R.A. (1995), A secondary structure model for the selfcleaving region of *Neurospora* VS RNA, *Proc. Natl. Acad. Sci.*, **92**, 4686-4690.

Beattie, T.L., and Collins, R.A. (1997), Identification of functional domains in the self-cleaving *Neurospora* VS ribozyme using damage selection, *J. Mol. Biol.*, **267**, 830-840.

Bertini, I. and Luchinat, C. (1986) NMR of paramagnetic molecules in biological systems, Menlo Park, California: Benjamin/Cummings.

Burkard, M.E., Kierzek, R. and Turner, D.H. (1999) Thermodynamics of unpaired terminal nucleotides on short RNA helixes correlates with stacking at helix termini in larger RNAs, *J. Mol. Biol.*, **290**, 967-982.

Brierly, I., Digard, P. & Inglis, S.C. (1989) Characterization of an efficient coronavirus ribosomal frameshifting signal: Requirement for an RNA pseudoknot, *Cell*, **57**, 537-547.

Brünger, A.T. (1992) *X-PLOR. A System for X-ray Crystallography and NMR.* Yale Univ. Press, New Haven.

Butcher, S.E., Allain, F.H.T, and Feigon, J. (1999) Solution structure of the loop B domain from the hairpin ribozyme, *Nature Struct. Biol.*, **6**, 212-216.

Butcher, S.E., Dieckmann, T. and Feigon, J. (1997) Solution structure of a GAAA tetraloop receptor RNA, *EMBO J.*, **16**, 7490-7499.

Cai, Z., Gorin, A. Frederick, R., Ye, X., Hu, W., Majumdar, A., Kettani, A. and Patel, D. (1998), Solution structure of P22 transcriptional antitermination N peptide-boxB RNA complex, *Nature Struct. Biol.*, **5**, 203–212.

Cai, Z. & Tinoco Jr., I. (1996), Solution structure of loop A from the hairpin ribozyme for Tobacco Ringspot virus satelite, *Biochemistry*, **35**, 6026-6036.

Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R. and Doudna, J.A. (1996a) Crystal structure of a Group I ribozyme domain: principles of RNA packing, *Science*, **273**,1678-1685.

Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Szewczak, A.A., Kundrot, C.E., Cech, T.R., Doudna, J.A. (1996b), RNA tertiary structure mediation by adenosine platforms, *Science*, **273**, 1696-1699.

Cate, J.H. and Doudna, J.A. (1996), Metal-binding sites in the major groove of a large ribozyme domain, *Structure*, **4**, 1221-1229.

Cate, J.H., Hanna, R.L. and Doudna, J.A. (1997), A magnesium ion core at the heart of a ribozyme domain, *Nat. Struct. Biol.*, **4**, 553-558.

Chamorro, M. Parkin, N. and Varmus, H.E. (1992), An RNA pseudoknot and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA, *Proc. Nat. Acad. Sci. USA*, **89**, 713–717.

Chen, X, Chamorro, M., Lee, S.I., Shen, L., Hines, J.V., Tinoco, I., Jr and Varmus, H.E. (1995), Structural and functional studies of retroviral RNA pseudoknots involved in ribosomal frameshifting: nucleotides at the junction of the two stems are important for efficient ribosomal frameshifting, *EMBO J.*, **14**, 842–852.

Chen, X., Kang, H., Shen, L.X., Chamorro, M., Varmus, H.E. and Tinoco, I. Jr. (1996), A characteristic bent conformation of RNA pseudoknots promotes -1 frameshifting during translation of retroviral RNA, *J. Mol. Biol.*, **260**, 479–483.

Cheong, C., Varani, G. and Tinoco, I.Jr (1990), Solution structure of an unusually stable RNA hairpin, 5'GGAC(UUCG)GUCC, *Nature*, **346**, 680-682.

Collins, R.A., and Olive, J.E. (1993), Reaction conditions and kinetics of self-cleavage of a ribozyme derived from *Neurospora* VS RNA, *Biochemistry*, **32**, 2795-2799.

Collins, R.A., and Saville, B.J. (1990) Independent transfer of mitochondrial chromosomes and plasmids during unstable vegetative fusion in *Neurospora*, *Nature*, **345**, 177-179.

Colmenarejo, G. and Tinoco I. Jr. (1999), Structure and thermodynamics of metal binding in the P5 helix of a Group I intron ribozyme, *J. Mol. Biol.*, **290**, 119-135.

Conn, G.L., Draper, D.E., Lattmann, E.E. & Gittis, A.G. (1999) Crystal structure of a conserved ribosomal protein-RNA complex, *Science*, **284**, 1171-1174.

Costa, M. and Michel, F. (1995), Frequent use of the same tertiary motif by self-folding RNAs, *EMBO J.*, **14**, 1276-1285.

Crothers, D.M., Cole, P.E., Hilbers, C.W. and Shulman, R.G. (1974), The molecular mechanism of thermal unfolding of Escherichia coli formylmethionine transfer RNA, *J. Mol. Biol.*, **87**, 63–88.

Deiman, B.A.L.M. and Pleij, C.W.A. (1997) Pseudoknots: a vital feature in viral RNA, Sem. Virol., 8, 166-175.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR.*, **6**, 277-293.

Dieckmann, T., Suzuki, E., Nakamura, G.K. and Feigon, J. (1996), Solution structure of an ATP-binding RNA aptamer reveals a novel fold, *RNA*, **2**, 628-40.

Dinman, J.D. & Wickner, R.B. (1992) Ribosomal frameshifting efficiency and gag/gag-pol ratio are critical for yeast M1 double stranded RNA virus propagation, *J. Virol*., **66**, 3669-3676.

Du, Z., Giedroc, D.P. and Hoffman, D.W. (1996), Structure of the autoregulatory pseudoknot within the gene 32 messenger RNA of bacteriophages T2 and T6: a model for a possible family of structurally related RNA pseudoknots, *Biochemistry*, **35**, 4187–4198.

Du, Z., & Hoffman, D.W. (1997) An NMR and mutational study of the pseudoknot within the gene 32 mRNA of bacteriophage T2: insights into a family of structurally related RNA pseudoknots, *Nucleic Acids Res*., **25**, 1130-1135.

Du, Z., Holland, J.A., Hansen, M.R., Giedroc.D.P. & Hoffman, D.W. (1997) Base pairings within the RNA pseudoknot associated with the simian retrovirus-1 gag-pro frameshift site, *J. Mol. Biol.*, **270**, 464-470.

Dumas, P., Mora, D., Florentz, C., Giegé, R., Verlaan, P., van Belkum, A. and Pleij, C.W.A. (1987), 3D- graphics modelling of the tRNA-like 3'-end of turnip yellow mosaic virus RNA: structural and functional implications, *J. Biomolec. Struct. and Dynam.*, **4**, 707–728.

Eckstein, F. and Lilley, D.M.J. (1996) Catalytic RNA, Springer-Verlag Berlin Heidelberg New York.

Farmer II, B.T., Muller, L., Nikonowicz, E.P. and Pardi, A. (1993a) Unambiguous resonance assignment in 13C,15N-labeled nucleic acids by 3D triple-resonance NMR, *J. Am. Chem. Soc.*, **115**, 11040-11041.

Farmer II, B.T., Muller, L., Nikonowicz, E.P. and Pardi, A. (1993b) Unambiguous through-bond sugar-to-base correlations for purines in 13C,15N-labeled nucleic acids: the HsCsNb,HsCs(N)bCb, and HbNbCb experiments, *J. Biomol. NMR*, **4**, 129-133.

Felden, B., Himeno, H., Muto, A., McCutcheon, J.P., Atkins, J.F. and Gesteland, R.F. (1997), Probing the structure of the Escherichia coli 10Sa RNA (tmRNA), *RNA*, **3**, 89–103.

Ferré–D'Amaré, A.R., Zhou, K. and Doudna, J.A. (1998), Crystal structure of a hepatitis delta virus ribozyme, *Nature*, **395**, 567–574.

Florentz, C., Briand, J.P., Romby, P., Hirth, L., Ebel, J.P. and Giegé, R. (1982), The tRNA-like structure of turnip yellow mosaic virus RNA: structural organization of the last 159 nucleotides from the 3'-OH terminus, *EMBO J.*, **1**, 269–276.

Freier, S.M., Burger, B.J., Alkema, D., Neilson, T. and Turner, D.H. (1983), Effects of 3'-dangling end stacking on the stabilit of GGCC and CCGG double helices, *Biochemistry*, **22**, 6198-6206.

Freier, S.M., Alkema, D., Sinclair, A., Neilson, T. and Turner, D.H. (1985), Contributions of dangling end stacking and terminal base pair formation to the stabilities of XGGCCp, XCCGGp, XGGCCYp, and XCCGGYp helixes, *Biochemistry*, **24**, 4533-4539.

Fuller, W., Hutchinson, F., Spencer, M. and Wilkins, M.H.F. (1967), Molecular and crystal structures of double-helical RNA, I. An X-ray diffraction study of fragmented yeast RNA and a preliminary double-helical RNA model, *J. Mol. Biol.*, **27**, 507-524.

Garcia, A., van Duin, J. and Pleij, C.W.A. (1993), Differential response to frameshift signals in eukaryotic and prokaryotic translational systems, *Nucleic Acids Res.*, **21**, 401–406.

Gautheret, D., Konings, D. & Gutell, R.R. (1994). A major family of motifs involving G.A mismatches in ribosomal RNA. *J. Mol. Biol*. **242**, 1-8.

Gesteland, R.F. & Atkins, J.F.(1996) Recoding: dynamic reprogramming of translation, *Annu. Rev. Biochem*., **65**, 741-768.

Gonzalez, R.L. Jr and Tinoco, I., Jr. (1999) Solution structure and thermodynamics of a divalent metal ion binding site in an RNA pseudoknot, *J. Mol. Biol.*, **289**, 1267-1282.

Giedroc, D.P., Theimer, C.A. & Nixon, P.L. (2000) Structure, stability and function of RNA pseudoknots involved in stimulating ribosomal frameshifting , *J Mol Biol.*, **298**, 167-85.

Griesinger, C., Otting, G., Wüthrich, K. and Ernst, R.R. (1988) Clean TOCSY for ^{1}H spin system identification in macromolecules, *J. Am. Chem. Soc.*, **110**, 7870-7872.

Grzesiek, S. and Bax, A (1995) Spin-locked multiple quantum coherence for signal enhancement in heteronuclear multidimensional NMR experiments, *J. Biomol. NMR*, **6**, 335-339.

Güntert, P., Mumenthaler, C. & Wüthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA, *J. Mol. Bio*l., **273**, 283-298.

Guo, H.C.T., De Abreu, D.M., Tillier, E.R.M., Saville, B.J., Olive, J.E., and Collins, R.A. (1993), Nucleotide sequence requirements for self-cleavage of *Neurospora* VS RNA, *J. Mol. Biol.,* **232**, 351-361.

Guo, H.C.T., and Collins, R.A. Collins (1995), Efficient trans-cleavage of a stem-loop RNA substrate by a ribozyme derived from *Neurospora* VS RNA, *EMBO J.*, **14**, 368-376.

Gutell, R.R., Cannone, J.J., Shang, Z., Du, Y. and Serra, M.J. (2000) A story: unpaired adenosine bases in ribosomal RNAs, *J. Mol. Biol.*, **304**, 335-354.

Haasnoot, C.A.G., Hilbers, C.W., van der Marel, G.A., van Boom, J.H., Singh, U.C., Pattabiraman, N. & Kollman, P.A., (1986) On loopfolding in nucleic acid hairpin-type structures, *J. Biom. Struct. and Dyn.*, **3**, 843-857

Hansen, M.R., Simorre, J.-P., Hanson, P., Mokler, V., Bellon, L., Beigelman, L. and Pardi, A. (1999) Identification and characterization of a novel high affinity metal-binding site in the hammerhead ribozyme, *RNA*, **5**, 1099-1104.

Hermann, T. and Patel, D.J. (1999), Stitching together RNA tertiary architectures, *J. Mol. Biol.*, **294**, 829-849.

Hermann, T. and Patel, D.J. (2000), Adaptive recognition by nucleic acid aptamers, *Science*, **287**, 820-825.

Heus, H.A. (1997), RNA aptamers, *Nat. Struct. Biol.*, **4**, 597-600.

Heus, H.A, van Kimmenade, J.M.A., van Knippenberg, P.H., Haasnoot, C.A.G., De Bruin, S.H. and Hilber, C.W. (1983) High resolution proton magnetic resonance studies of the 3'-terminal Colicin fragment of 16S ribosomal RNA from Escherichia coli, *J. Mol. Biol., **170,*** 939-956.

Heus, H.A. and Pardi, A. (1991) Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops, *Science*, **253**, 191-194.

Heus, H.A., Wijmenga, S.S., Hoppe, H., Hilbers, C.W. (1997), The detailed structure of tandem G.A mismatched base-pair motifs in RNA duplexes is context dependent , *J. Mol. Biol.*, **271**, 147-158.

Hilbers, C.W., Haasnoot, C.A.G. de Bruin, S.H., Joordens, J.J.M., van der Marel, G.A. and van Boom, J.H. (1985), Hairpin formation in synthetic oligonucleotides, *Biochimie*, **67**, 685-695.

Hilbers, C.W., Heus, H.A., van Dongen, M.J.P. and Wijmenga, S.S. (1994) In: *Nucleic Acids and molecular biology* (Eckstein, F. and Liley, D., Eds), **8**, 56–104.

Hilbers, C.W., Michiels, P.J.A. & Heus, H.A. (1998) New developments in structure determination of pseudoknots, *Biopolymers*, **48**, 137-153.

Hoogstraten, C.G., Legault, P., Pardi, A. (1998), NMR solution structure of the lead-dependent ribozyme: evidence for dynamics in RNA catalysis, *J. Mol. Biol*., **284**, 337-350.

Ikura, M., Kay, L.E., Tschudin, R. & Bax, A. (1990) Three-dimensional NOESY-HMQC spectroscopy of a ^{13}C labeld protein, *J. Magn. Reson.*, **86**, 204-209.

Ippel, J.H., Wijmenga, S., de Jong, R., Heus, H.A., Hilbers, C.W., de Vroom, E., van der Marel, G.A. and van Boom, J.H. (1996) Heteronuclear scalar couplings in the bases and sugar rings of nucleic acids: Their determination and application in assignment and conformational analysis, *Magn. Res. Chem.*, **34**, S156.

Ippolito, J.A. and Steitz, T.A. (2000), The structure of the HIV-1 RRE high affinity rev binding site at 1.6 Å resolution, *J. Mol. Biol.*, **295**, 711-717.

Jaeger, J., Restle, T. and Steitz, T.A. (1998), The structure of HIV-1 reverse transcriptase complexed with an RNA pseudoknot inhibitor, *EMBO J.*, **17**, 4535–4542.

Jeener, J., Meier, B.H., Bachmann, P. and Ernst, R.R. (1979) Investigation of exchange processes by two-dimensional NMR spectroscopy, *J. Chem. Phys.*, **71**, 4546-4553.

Jentsch, S. (1996), When proteins receive deadly messages at birth, *Science*, **271**, 955–996.

Jiang, F., Kumar, R.A., Jones, R.A. and Patel, D.J. (1996) Structural basis of RNA folding and recognition in an AMP-RNA aptamer complex, *Nature*, **382**, 183-186.

Jucker, F.M., Heus, H.A., Yip, P.F., Moors E.H., and Pardi, A. (1996) A network of heterogeneous hydrogen bonds in GNRA tetraloops, *J. Mol. Biol.*, **264**, 968-980.

Jucker, F.M. and Pardi, A. (1995), GNRA tetraloops make a U-turn, *RNA*, **1**, 219-222.

Kalurachchi, K. and Nikonowicz, E.P. (1998), NMR structure determination of the binding site for ribosomal protein S8 from Escherichia coli 16 S rRNA, *J. Mol. Biol.*, **280**, 639-654.

Kang, H. Hines, J.V. and Tinoco, I. Jr. (1996), Conformation of a non-frameshifting RNA pseudoknot from mouse mammary tumor virus, *J. Mol. Biol.*, **259**, 135–147.

Kang, H. and Tinoco, I. Jr. (1997), A mutant RNA pseudoknot that promotes ribosomal frameshifting in mouse mammary tumor virus, *Nucleic Acids Res.*, **25**, 1943–1949.

Keiler, K.C., Waller, P.R.H. and Sauer, R.T. (1996), Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA, *Science*, **271**, 990–993.

Kieft, J.S. and Tinoco, I. Jr. (1995), Solution structure of a metal-binding site in the major groove of RNA complexed with cobalt (III) hexammine, *Structure*, **5**,713-721.

Kellog, G.W. & Schweizer, B.I. (1993) Two- and three-dimensional 31P-driven NMR procedures for complete assignment of backbone resonances in oligodeozynucleotides, *J. Biomol. NMR*, **3**, 577-595.

Kim, S.H., Quigley, G.J., Suddath, F.L., McPherson, Sneden, D., Kim, J.J., Weinzierl, J. and Rich, A. (1973), Three-dimensional structure of yeast phenylalanine transfer RNA: folding of the polynucleotide chain, *Science*, **179**, 285-288.

Kim, Y.-G., Su, L., O'Neill, A. & Rich, A. (1999) Specific mutations in a viral RNA pseudoknot drastically change ribosomal frameshifting efficiency, *Proc. Natl. Acad. Sci.*, **96**, 14234-14239.

Kim, S.H., Suddath, F.L., Quigley, G.J., McPherson, A., Sussman, J.L., Wang, A.H., Seeman, N.C. and Rich, A. (1974), Three-dimensional tertiary structure of yeast phenylalanine transfer RNA, *Science*, **185**, 435-440.

Kolk, M.H., Heus, H.A. and Hilbers, C.W. (1997), The structure of the isolated, central hairpin of the HDV antigenomic ribozyme: novel structural features and similarity of the loop in the ribozyme and free in solution, *EMBO J.*, **16**,3685-3692.

Kolk, M.H., van der Graaf, M., Wijmenga, S.S., Pleij, C.W., Heus, H.A. and Hilbers, C.W. (1998a), NMR structure of a classical pseudoknot: interplay of single- and double-stranded RNA, *Science*, **17**, 434-438.

Kolk, M.H., van der Graaf, M., Fransen, C.T., Wijmenga, S.S., Pleij, C.W., Heus, H.A. and Hilbers, C.W. (1998b), Structure of the 3'-hairpin of the TYMV pseudoknot: preformation in RNA folding, *EMBO J.*, **17**, 7498-7504.

Koradi, R., Billeter, M. and Wüthrich, K., (1996) MOLMOL: a program for display and analysis of macromolecular structures, J. *Mol. Graphics*, **14**, 51-55.

Larsen, B., Gesteland, R.F. & Atkins, J.F. (1997) Structural probing and mutagenic analysis of the stem-loop required for Escherichia coli dnaX ribosomal frameshifting: Programmed frameshifting of 50%, *J. Mol. Biol*., **271**, 47-60.

Larsen, B., Wills, N.M., Gesteland, R.F & Atkins, J.F. (1994) rRNA-mRNA base pairing stimulates a programmed –1 ribosomal frameshift, *J. Bacteriol.*, **176**, 6842-6851.

Laughlan, G., Murchie, A.I.H., Norman, D.G., Moore, M.H., Moody, P.C.E., Lilly, D.M.J. and Louisi, B. (1994), The high-resolution crystal structure of a parallel-stranded guanine tetraplex, *Science*, **265**, 520-524.

Lavery, R. & Sklenár, H. (1988) The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. struct. Dynam.* **6,** 63-91.

Lavery, R. & Sklenár, H. (1989) Defining the structure of irregular nucleic acids: conventions and principles, *J. Biomol. struct. Dynam.* **6**, 655-667.

Legault, P., Pardi, A. (1994), In situ probing of adenine protonation in RNA by ^{13}C NMR, *J. Am. Chem. Soc.*, **116**, 8390-8391.

Legault, P., Farmer II, B.T., Mueller, L. and Pardi, A. (1995) Through-bond correlation of adenine protons in a 13C-labeled ribozyme, *J. Am. Chem. Soc.*, **117**, 11043-11048.

Legault, P., Li, J., Morgridge, J., Lay, L.E. and Greenblatt, J. (1998), NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif, *Cell*, **93**, 289–299.

Liphardt, J., Napthine, S., Kontos, H. & Brierly, I. (1999) Evidence for an RNA pseudoknot loop-helix interaction essential for efficient –1 ribosomal frameshifting, *J. Mol. Biol.*, **288**, 321-335.

Lippens, G., Dhalluin, C. & Wieruszeski, J.-M. (1995) Use of a water flip-back pulse in the homonuclear NOESY experiment, *J. Biomol. NMR*, **5**, 327-331.

Lynch, S.R. and Puglisi, J.D. (2000) Application of Residual Dipolar Coupling Measurements To Identify Conformational Changes in RNA Induced by Antibiotics , *J. Am. Chem. Soc.*, **122**, 7853-7854.

Marino, J.P., Prestegard, J.H. and Crothers, D.M. (1994) Correlation of adenine H2/H8 resonances in uniformly labeled ^{13}C labeled RNAs by 2D HCCH-TOCSY: a new tool for ^{1}H assignment, *J. Am. Chem. Soc.*, **116**, 2205-2206.

Marino, J.P., Diener, J.L., Moore, P.B. and Griesinger, C. (1997), Multiple-quantum coherence dramatically enhances the sensitivity of CH and CH2 correlations in uniformly 13C-labeled RNA, *J. Am. Chem. Soc.*, **119**, 7361-7366

Michiels, P.J.A., Versleijen, A.A.M., Verlaan, P.W., Pleij ^{Y}, C.W.A., Hilbers, C.W. and Heus, H.A. (2001), Solution structure of the pseudoknot of SRV-1 RNA, involved in ribosomal frameshifting, *Chapter 5 in this thesis*.

Milligan, J.F., Groebe, D.R., Witherell, G.W. and Uhlenbeck, O.C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates, *Nucleic Acids Res.*, **15**, 8783-8798.

Mizuno, M. and Sundaralingam, M. (1978) Stacking of Crick Wobble pair and Watson-Crick pair: stability rules of G-U pairs at ends of helical stems in tRNAs and the relation to codon-anticodon Wobble interaction, *Nucleic Acids Res.*, **6,** 4451-4461.

Mollova, E.T., Hansen, M.R. and Pardi, A. (2000) Global Structure of RNA Determined with Residual Dipolar Couplings, *J. Am. Chem. Soc.*, **122**,11561-11562.

Nameki, N., Felden, B., Atkins, J.F., Gesteland, R.F., Himeno, H. and Muto, A. (1999), Functional and structural analysis of a pseudoknot upstream of the tag-encoded sequence in E. coli tmRNA, *J. Mol. Biol.*, **286**, 733–744.

Nissen, P., Ippolito, J.A., Ban, N., Moore, P.B. and Steitz, T. A. (2001), RNA tertiary interactions in the large ribosomal subunit: The A-minor motif, *Proc. Natl. Acad. Sci. USA*, **98**, 4899-4903.

Pan, B., Mitra, S.N., Sundaralingam, M. (1998), Structure of a 16-mer RNA duplex r(GCAGACUUAAAUCUCGC)$_2$ with wobble C.A$^+$ mismatches, *J. Mol. Biol.*, **283**, 977-984.

Parkin, N.T., Chamorro, M. & Varmus, H.E. (1992) Human immunodeficiency virus type 1 gag-pol frameshifting is dependent on downstream mRNA secondary structure: Demonstration by expression in vivo, *J. Virol.*, **66**, 5147-5151.

Perrotta, A.T. and Been M.D. (1991), A pseudoknot-like structure required for efficient self-cleavage of hepatitis delta virus RNA, *Nature*, **350**, 434–436.

Philippe C., Eyermann, F., Bénard, L., Portier, C., Ehresmann, B. and Ehresmann, C. (1993) Ribosomal protein S15 from Escherichia coli modulates its own translation by trapping the ribosome on the mRNA initiation loading site, *Proc. Natl. Acad. Sci. USA*, **90**, 4394-4398.

Pinck, M., Yot, P., Chapeville, F. and Duranton, H. (1970), Enzymatic binding of valine to the 3' end of TYMV-RNA, *Nature*, **226**, 954 – 956.

Plateau, P. and Guéron, M. (1982) Exchangeable proton NMR without base-line distortion, using new strong-pulse sequences, *J. Am. Chem. Soc.*, **104**, 7310-7311.

Pleij, C.W.A. (1990), Pseudoknots: a new motif in the RNA game, *Trends Biochem. Sci.*, **15**, 143–147.

Pleij, C.W.A. (1994), RNA pseudoknots, *Current Opinion Struct. Bio*., **4**, 337–344.

Pleij, C.W., Rietveld, K. and Bosch, L. (1985) A new principle of RNA folding based on pseudoknotting, *Nucleic Acids Res.*, **13**, 1717-1731.

Pley, H.W., Flaherty, K.M. and McKay, D.B. (1994a), Three-dimensional structure of a hammerhead ribozyme, *Nature*, **372**, 68-74.

Pley, H.W., Flaherty, K.M. and McKay, D.B. (1994b), Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix, *Nature*, **372**,111-113.

Power, M.D., Marx, P.A., Bryant, M.L., Gardner, M.B., Barr, P.J. & Luciw, P.A. (1986) Nucleotide sequence of SRV-1, a type D simian acquired Immune deficiency syndrome retrovirus, *Science*, **231**, 1567-1572.

Puglisi, J.D., Tan, R., Calnan, B.J., Frankel, A.D. and Williamson, J.R. (1992), Conformation of the TAR RNA-arginine complex by NMR spectroscopy, *Science*, **257**, 76-80.

Puglisi, J.D., Wyatt, J.R., and Tinoco, I. Jr. (1990a), Conformation of an RNA pseudoknot, *J. Mol. Biol.*, **214**, 437–453.

Puglisi, J.D., Wyatt, J.R., and Tinoco Jr., I. (1990b), Solution conformation of an RNA hairpin loop, *Biochemistry*, **29**, 4215-4226.

Rastogi, T., Beattie, T.L., Olive, J.E., and Collins, R.A. (1996), A long-range pseudoknot is required for activity of the *Neurospora* VS ribozyme, *EMBO J.*, **15**, 2820-2825.

Rastogi, T. and Collins, R.A. (1998) Smaller, faster ribozymes reveal the catalytic core of *Neurospora* VS RNA, *J. Mol. Biol.*, **277**, 215-224.

Rettberg, C.C., Prère, M.F., Gesteland, R.F., Atkins, J.F. & Fayet, O. (1999) A three-way junction and constituent stem-loops as the stimulator for -1 frameshifting in bacterial insertion sequence IS911, *J. Mol. Biol*., **286**, 1365-1378.

Rich, A. and Watson, J.D. (1954), *Proc.Nat. Acad.Sci*., **40**, 759

Riek, R., Pervushin, K., Fernandez, C., Kainosho, M and Wütrich, K. (2001) [13C,13C]- and [13C,1H]-TROSY in a triple resonance experiment for ribose-base and intrabase correlations in nucleic acids, *J. Am. Chem. Soc.*, **123**, 658-664.

Rietveld K., Van Poelgeest, R., Pleij, C.W., Van Boom, J.H. and Bosch, L. (1982), The tRNA-like structure at the 3' terminus of turnip yellow mosaic virus RNA. Differences and similarities with canonical tRNA, *Nucleic Acids Res.*, **10**, 1929-1946.

Roberts, R.W. and Crothers, D.M. (1996), Kinetic discrimination in the folding of intramolecular triple helices, *J. Mol. Biol.*, **260**, 135–146.

Robertus, J.D., Ladner, J.E., Finch, J.T., Rhodes, D., Brown, R.S., Clark, B.F. and Klug, A. (1974), Structure of yeast phenylalanine tRNA at 3 Å resolution, *Nature*, **250**, 546-551.

Rudisser S. and Tinoco I Jr. (2000), Solution structure of Cobalt(III)hexammine complexed to the GAAA tetraloop, and metal-ion binding to G.A mismatches, *J. Mol. Biol.*, **295**,1211-1223.

Saenger, W. (1984) Principles of nucleic acid structure, Springer-Verlag, New-York.

Santa Lucia Jr., J. & Turner, D.H. (1993), Structure of $(rGGCGAGCC)_2$ in solution from NMR and restrained molecular dynamics, *Biochemistry*, **32**, 6026-6036.

Saville, B.J., and Collins, R.A. (1990), A site-specific self-cleavage reaction performed by a novel RNA in *Neurospora* mitochondria, *Cell*, **61**, 685-696.

Saville, B.J., and Collins, R.A. (1991), RNA-mediated ligation of self-cleavage products of a *Neurospora* mitochondrial plasmid transcript, *Proc. Natl. Acad. Sci.*, **88**, 8826-8830.

Schindelin, H., Zhang, M., Bald, R., Furste, J. P., Erdmann, V. A. & Heinemann, U. (1995) Crystal structure of an RNA dodecamer containing the Escherichia coli Shine- Dalgarno sequence, *J. Mol. Biol*., **249**, 595-603.

Schweizer, M. P., Broom, A.D., Ts'o, P.O.P. and Hollis, D. (1968), Studies of inter- and intramolecular interactions in mononucleotides by proton magnetic resonance, *J. Amer. Chem. Soc.*, **90**, 1042-1055.

Scott, W.G., Finch, J.T., and Klug, A. (1995), The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage, *Cell*, **81**, 991-1002.

Sen, D., and Gilbert, W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis, Nature 334, 364-366.

Serra, M.J., Lyttle, M.H., Axenson, T.J., Schadt, C.A. and Turner, D.H. (1993), RNA hairpin loop stability depends on closing base pair, *Nucleic Acids Res.*, **21**, 3845–3849.

Shaka, A.J. and Freeman, R. (1983) Simplification of NMR spectra by filtration through multiple-quantum coherence, *J. Magn. Reson*., **51**, 169-173.

Shen, L.X. and Tinoco, I. Jr. (1995), The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus, *J. Mol. Biol.*, **247**, 963-978.

Simorre, J.P., Zimmermann, G.R., Mueller, L. and Pardi, A. (1996) Correlation of the guanosine exchangeable and nonexchangeable base protons in 13C-/15N-labeled RNA with an HNC-TOCSY-CH experiment, *J. Biomol. NMR*, **7**, 153-156.

Sklenár, V., Dieckmann, T., Butcher, S.E. and Feigon, J. (1998) Optimization of Triple-Resonance HCN Experiments for Application to Larger RNA Oligonucleotides, *J. Mag. Res.,* **130**, 118-124.

Sklenár, V., Miyashiro, H., Zon, G., Miles, H.T. and Bax, A. (1986) Assignment of the 31P and 1H resonances in oligonucleotides by two-dimensional NMR spectroscopy, *FEBS Lett.*, **208**, 94-98.

Sklenár, V., Peterson, R.D., Rejante, M.R. and Feigon, J. (1993a) Two- and three-dimensional HCN experiments for correlating base and sugar resonances in 15N ,13C-labeled RNA oligonucleotides, *J. Biomol. NMR*, **3**, 721-727.

Sklenár, V., Peterson, R.D., Rejante, M.R. and Feigon, J. (1993b) Two-dimensional triple-resonance HCNCH experiment for direct correlation of ribose H1' and base H8, H6 protons in 13C ,15N-labeled RNA oligonucleotides, *J. Am. Chem. Soc.*, **115**, 12181-12182.

Sklenár, V., Piotto, M., Leppik, R. and Saudek, V. (1993c), Gradient-tailored water suppression for 1H-15N HSQC experiments optimized to retain full sensitivity, *J. Magn. Reson. A*, **102**, 241-245.

Smith, F.W. and Feigon, J.(1992), Quadruplex structure of Oxytricha telomeric DNA oligonucleotides, *Nature*, **356**, 164-168.

Solie, T.N. and Schellman, J.A. (1968), The interaction of nucleosides in aqueous solution, *J. Mol. Biol.*, **33**, 61-77.

Somogyi, P., Jenner, A.J., Brierley, I. & Inglis, S.C. (1993) Ribosomal pausing during translation of an RNA pseudoknot, *Mol. Cell. Biol.*, **13**, 6931-6940.

Sood, V.D., Beattie, T.L., and Collins, R.A. (1998), Identification of phosphate groups involved in metal binding and tertiairy interactions in the core of the *Neurospora* VS RNA, *J. Mol. Biol.*, **282**, 741-750.

Spedding, G., Gluick, T.C. and Draper, D.E. (1993) Allosteric mechanism for translational repression in the Escherichia coli alpha operon, *Proc. Natl. Acad. Sci. USA*, **90**, 4399-4403.

Stein, E.G., Rice, L.M. and Brünger, A.T. (1997) Torsion-angle dynamics as a new efficient tool for NMR structure calculation. *J. Magn. Reson.*, **124**, 154-164.

Su, L., Chen, L., Egli, M., Berger, J.M. and Rich, A. (1999), Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot, *Nature Struct. Biol.*, **6**, 285–292.

Sung, D. and Kang, H. (1998), Mutational analysis of the RNA pseudoknot involved in efficient ribosomal frameshifting in simian retrovirus-1, *Nucleic Acids Res.*, **26**, 1369–1372.

Szewczak, A.A., Moore, P.B., Chang, Y.L., Wool, I.G. (1993), The conformation of the sarcin/ricin loop from 28S ribosomal RNA, *Proc. Natl. Acad. Sci.*, **90**, 9581-9585.

Tang, C.K. and Draper, D.E. (1989), Unusual mRNA pseudoknot structure is recognized by a protein translational, *Cell*, **57**, 531–536.

Tate, S-I, Ono, A. and Kainosho, M. (1994) An alternative triple-resonance method for the through-bond correlation of intranucleotide H1' and H8 NMR signals of purine nucleotides. Application to a DNA dodecamer with fully 13C/15N-labeled deoxyadenosine residues, *J. Am. Chem. Soc.*, **116**, 5977-5978.

Ten Dam, E.B., Brierly, I., Ingles, S. & Pleij, C.W.A. (1994) Identification and analysis of the pseudoknot containing gag-pro ribosomal frameshift signal of simian retrovirus-1, *Nucleic Acids Res.*, **22**, 2304-2310.

Ten Dam, E.B., Pleij, C.W.A. and Draper. D.E. (1992), Structural and functional aspects of RNA pseudoknots, *Biochemistry*, **31**, 11665–11676.

Ten Dam, E.B. Pleij, C.W.A. & Bosch L. (1990) RNA pseudoknots: Translational frameshifting and readthrough on viral RNAs. *Virus Genes*, **4**, 121-136.

Ten Dam, E.B., Verlaan, P.W.G and Pleij, C.W.A. (1995) Analysis of the role of the pseudoknot in the SRV-1 gag-pro ribosomal frameshift signal: loop lengths and the stability of the stem regions, *RNA*, **1**, 146-154.

Tougad, P., Chantot, J-F. and Guschlbauer, W. (1973), Nucleoside conformations X., An X-ray fiber diffraction study of the gels of guanine nucleosides, *Biochim. Biophys. Acta*, **308**, 9-16.

Trantirék, L., Urbášek, M., Štelf, R., Feigon, J. and Sklenář, V. (2000), A method for direct determination of helical parameters in nucleic acids using residual dipolar couplings, *J. Am. Chem. Soc.*, **122**, 10454-10455.

Ts'o, P.O.P. and Chan S.I. (1964), Interaction and association of bases and nucleosides in Aqueous solutions. II. Association of 6-methylpurine and 5-bromouridine and treatment of multiple equilibria, *J. Amer. Chem. Soc.*, **86**, 4176-4181.

Tsuchihashi, Z. (1991) Translational frameshifting in the Escherichia coli dnaX gene in vitro, *Nucleic Acids Res.*, **19**, 2457-2462.

Tuerk, C. and Gold, L. (1990), Systematic Evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase, *Science*, **249**, 505–510.

Tzeng, T.-H., Tu, C.-L. & Bruenn, J.A. (1992) Ribosomal frameshifting requires a pseudoknot in the Saccharomyces cerevisiae double-stranded RNA virus, *J. Virol.*, **66**, 999-1006.

Van Belkum, A., Wiersema, P.J., Joordens, J., Pleij, C.W.A., Hilbers, C.W. and Bosch, L. (1989), Biochemical and biophysical analysis of pseudoknot-containing RNA fragments. Melting studies and NMR spectroscopy, *Eur. J. Biochem.*, **183**, 591–601.

Van Dongen, M.J., Wijmenga, S.S., Eritja, R., Azorin, F. and Hilbers, C.W. (1996) Through-bond correlation of adenine H2 and H8 protons in unlabeled DNA fragments by HMBC spectroscopy, *J. Biomol. NMR*, **8**, 207-212.

Watson, J.D. (1999), Prologue to the first edition. Early speculations and facts about RNA templates, In Gesteland, R.F., Cech, T.R., Atkins, J.F. (eds), *The RNA world; the second edition*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. xvii-xxv.

Watson, J.D. and Crick, F.H.C. (1953), A structure for deoxyribose nucleic acid, *Nature*, **171**, 737-738.

Walter, A.E., Wu, M. and Turner, D.H. (1994) The stability and structure of tandem GA mismatches in RNA depend on closing base pairs, *Biochemistry*, **33**, 11349-11354.

Wedekind, J.E., and McKay, D.B. (1999), Crystal structure of a lead-dependent ribozyme revealing metal binding sites revelant to catalysis, *Nature Struct. Biol.*, **6**, 261-268.

Westhof, E., Masquida, B. and Jaeger, L. (1996) *Folding and Design* **1**, R78–R88.

Wijmenga, S.S. & van Buuren, B.N.M. (1998) The use of NMR methods for conformational studies of nucleic acids, *Prog. Nucl. Magn. Res. Spec.*, **32**, 287-387.

Wijmenga, S.S., Mooren, M.W. and Hilbers, C.W. (1993) NMR of nucleic acids; from spectrum to structure. In Roberts, G.C.K. (eds), *NMR of macromolecules. A practical approach*. Oxford University Press, New York, pp. 217-288.

Williamson, J. R., Raghuraman, M. K. and Cech, T. R. (1989). Monovalent cation-induced structure of telomeric DNA: the G-quartet model, *Cell* **59**, 871-880.

Wimberly, B.T., Guymon, R., McCutcheon, J.P., White, S.W. and Rmakrishnan, V. (1999), A detailed view of a ribosomal active site: the structure of the L11-RNA complex, *Cell*, **97**,491-502.

Woese, C.R., Winker, S. and Gutell, R.R. (1990), Architecture of ribosomal RNA: constraints on the sequence of "tetra-loops", *Proc. Natl. Acad. Sci. USA*, **87**, 8467-8471.

Wyatt, J.R. and Tinoco, I., Jr. (1993), RNA structural elements and RNA function. In the RNA world (Gesteland, R.F. and Atkins, J.F., eds.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 465-496.

Ye, X., Gorin, A., Ellington, A.D., Patel, D.J. (1996), Deep penetration of an alpha-helix into a widened RNA major groove in the HIV-1 rev peptide-RNA aptamer complex, *Nat. Struct. Biol.*, **3**, 1026-1033.

Zimmermann, G.R., Jenison, R.D., Wick, C.L., Simorre, J.P. and Pardi, A. (1997) Interlocking structural motifs mediate molecular discrimination by a theophylline-binding RNA, *Nat. Struct. Biol.*, **4**, 644-649.

Appendices

***Table A.I**. Proton and phosphorus chemical shifts of the VSm1 hairpin (at pH 5.0) relative to TSP.*

	H8/H6	H2/H5	H1'	H2'	H3'	H4'	^{31}P	im/am[c]
G1	8.06	n.a.[a]	5.76	4.86	4.46	4.00		
U2	7.85	5.70	5.90	4.44	4.45	4.55		11.90
G3	8.03	n.a.	5.77	4.60	4.56	4.49	-2.34	13.20
C4	7.33	5.00	5.45	4.52	4.27	4.43	-2.40	8.44/6.76[c]
G5	8.00	n.a.	5.63	4.48	4.83	4.61	-2.56	9.77 6.85/5.90[c]
A6	8.07	7.94	5.32	4.78	4.54	4.46[b]	-1.85	
A7	8.02	8.11	5.22	4.57	4.54	4.46	-2.89	
G8	6.87	n.a.	5.41	4.47	4.36	4.43	-2.04	12.38
A9	7.76	7.68	5.97	4.46	4.59	4.37[b]	-2.71	
C10	7.17	5.02	5.45	4.37	4.31	4.46	-2.23	8.19/6.82[c]
G11	7.56	n.a.	5.70	4.42	4.66		-1.99[b]	10.75
A12	8.36	7.82	5.69	4.75	4.46	4.22	0.25	
A13	7.96	7.80	5.43	4.37	4.59	4.30	-1.26	
A14	8.22	8.14	6.00	4.65	5.05	4.46	-3.10	
G15	7.89	n.a.	**3.62**	4.22	4.12	4.30[b]	-0.78	13.18
U16	7.76	5.04	5.56	4.51	4.43	4.45[b]	-3.07	14.45
C17	7.69	5.56	5.43	4.47	4.31		-2.33	8.14/6.91[c]
C18	7.67	5.64	5.45	3.93	4.36		-2.49	
G19	8.14	n.a.	5.53	4.76	4.91	4.36[b]	-3.15	9.98
A20	7.58	8.03	5.58	4.80	4.40	4.49	-0.08	
G21	7.53	n.a.	**4.15**	4.22	4.29		-2.22	12.91
C22	7.39	5.20	5.49	4.52	4.35	4.42	-2.60	8.00/6.49[c]
G23	7.54	n.a.	5.72	4.49	4.36	4.42	-1.98	11.27
C24	7.58	5.40	5.64	4.00	4.12	4.14	-2.71	

[a]Not applicable

[b]Tentative assignment.

[c]Amino protons

Table A.II. *Structural statistics of the final ensemble of 20 structures.*

Distance restraints	**all**	**internal loop***
Intranucleotide	120	52
Internucleotide	173	56
Non-NOE	5	5
Rms deviations (Å, °)		
Distance restraints (298)	0.071±0.002	0.064±0.004
Dihedral restraints (134)	0.960±0.079	1.064±0.136
Rms deviation from idealized geometry		
Bonds (Å)	0.0111±0.0003	0.0111±0.0003
Angles (°)	1.948±0.075	1.944±0.075
Impropers (°)	0.510±0.022	0.333±0.037
Restraint violations		
Number of distance violations† > 0.3Å	2±1	0±1
Number of dihedral violations‡ > 3°	5±3	3±3
Atomic rms deviations (Å)¶	1.65	0.77

*Residues 4 through 8 and 17 through 21. †None larger than 0.5Å. ‡None larger than 5°. ¶With respect to the average structure.

Table A.III*. Carbon chemical shifts of all adenines from the RNA pseudoknot of SRV-1 relative to TSP.*

	C2	C4	C5	C6	C8
A7	154.4	151.9	121.2	157.4	142.8
A13	152.6	148.8	120.7	157.1	138.9
A21	156.6	150.8	121.0	158.2	141.1
A22	155.2	150.9	120.5	157.5	141.1
A23	155.0	151.1	120.5	157.1	141.6
A25	154.5	149.6	120.7	156.4	141.7
A26	155.5	149.1	120.5	157.2	140.4
A28	156.6	152.1	120.9	157.7	142.4
A32	153.7	148.6	120.8	157.8	138.8
A35	154.3	149.7	121.4	157.8	139.3

***Table A.IV**. Proton and phosphorus chemical shifts of the RNA pseudoknot of SRV-1 relative to TSP.*

	H8/H6	H2/H5	H1'	H2'	H3'	H4'	H5'/"	^{31}P	im & am[c]
G1	8.19	na	5.77	4.86	4.81	4.56	4.29/4.42		13.28
C2	7.88	5.46	5.82	4.86	4.62	4.56	4.37/4.55		6.73/8.62[c]
G3	7.61	na	5.80	4.66	4.50		4.26/4.50	-1.75	12.69
G4	7.44	na	5.10	4.46	4.42				13.77 9.39[c]
C5	7.40	5.23	5.93	4.34	4.46	4.48	4.13/4.48[b]		6.91/8.42[c]
C6	7.86	5.52	5.82	4.23	4.34	4.19[b]	4.18/4.81[b]	-3.19	
A7	8.54	8.34	6.21	5.04	4.96	4.65	4.29/4.31		
G8	8.20	na	5.90	4.80	4.75	4.59		-1.54	12.61
C9	8.05	5.22	5.60	4.33	4.63	4.49	4.16/4.64	-3.43	6.75/8.53[c]
U10	7.97	5.36	5.57	4.41	4.59	4.45	4.14/4.57	-2.52	14.11
C11	7.92	5.62	5.60	4.24	4.50	4.44	4.13/4.54	-2.43	6.93/8.36[c]
C12	7.78	5.46	5.45	4.58	4.54	4.39	4.13/4.54		6.82/8.20[c]
A13	7.77	7.23	5.89	4.76	4.47	4.54	4.18/4.49		
G14	7.08	na	5.11	4.61	4.42		4.08/4.45		12.76
G15	6.99	na	5.10	4.38	4.20	4.49	3.98/4.37		13.70 9.46[c]
C16	7.63	5.02	5.51	4.76	4.61	4.46		-3.00	6.74/8.68[c]
C17	7.75	5.68	5.65	4.89	4.54	4.77	4.24/4.62	-1.39	6.86/8.55[c]
G18	7.80	na	5.91	4.59	4.64	4.42			12.86
C19	7.62	5.42	5.97	4.16	4.54	4.01	3.93/4.49	-2.77	
C20	7.56	6.07	5.74	4.15	4.42	3.78	3.54/3.64	-1.23	
A21	7.94	8.41	5.58	4.42	4.75	4.31	3.91/4.10	-1.21	
A22	8.20	7.54	5.45	4.61	4.76	4.50	4.16/4.27	-1.86	
A23	8.29	7.81	5.80	4.68	4.83	4.53	4.26/4.28	-1.77	
C24	7.79	5.91	5.93	4.51	4.67	4.55	4.29/4.48	-1.65	
A25	8.29	7.48	5.78	4.76	4.64[b]	4.64	4.33/4.51	-1.51	
A26	8.32	8.40	5.96	4.37	4.87	4.61			
U27	7.26	5.01	5.87	4.11	4.52	4.30	4.08/4.44	-1.90	
A28	8.40	8.23	6.19	4.78	4.97	4.73	4.20/4.23		
U29	8.00	5.70	5.81	4.70	4.59	4.53	4.32/4.34	-1.76	13.86
G30	8.04	na	5.92	4.63	4.75	4.55			11.97
G31	7.39	na	5.69	4.43	4.68	4.48		-2.03	12.35
A32	7.73	7.51	5.90	4.54	4.68	4.47	4.15/4.57		
G33	7.03	na	5.52	4.38	4.29	4.42	4.04/4.41		13.55
C34	7.19	5.19	5.36	4.52	4.30	4.37	4.02/4.40		6.88/8.12[c]
A35	7.95	7.41	6.00	4.33	4.52	4.44	4.10/4.51		
C36	7.49	5.42	5.52	3.92	4.10	4.12	4.01/4.32	-2.40	

[a]Not applicable

[b]Tentative assignment.

[c]Amino protons

***Table A.V**. Carbon chemical shifts of the RNA pseudoknot of SRV-1 relative to TSP.*

	C2/C5	C8/C6	C1'	C2'	C3'	C4'	C5'
G1	na	139.1	91.4			83.9	
C2	97.8	141.7	92.8			82.4	66.3
G3	na	136.2	93.0				
G4	na	136.2					
C5	96.5	140.5	94.2			82.9	64.6
C6	98.0	141.8	93.7			82.4	65.1
A7	154.1	142.9	88.7	76.2	78.4	86.6	68.4
G8	na	139.1	94.0			83.1	
C9	97.1	142.2	94.2			82.3	64.3
U10	103.3	142.3	93.6	75.4	72.0	82.0	64.2
C11	97.5	141.6	93.6			81.8	64.7
C12	97.9	140.8	93.4			81.7	64.7
A13	152.6	138.8	93.1	75.6	73.0	82.0	66.1
G14	na	135.5	92.8				65.0
G15	na	135.6	92.8				66.3
C16	96.0	141.0	94.6			82.6	
C17	98.7	140.8	94.9			82.3	65.8
G18	na	136.8	93.0				
C19	98.7	142.1	90.2			84.6	66.9
C20	99.6	143.5	89.9			83.7	67.3
A21	156.6	141.1	89.8	74.9	76.0	83.5	67.8
A22	155.1	141.2	89.0	75.3	76.5	84.9	67.7
A23	155.0	141.6	88.5	75.7	76.1	84.7	
C24	98.8	143.9	93.0	76.1	75.6	84.2	67.7
A25	154.5	141.7	93.6		75.9	84.1	67.6
A26	155.5	140.4	92.3	76.3	72.2	82.7	
U27	103.6	140.9	92.9	76.2	73.3	82.1	65.2
A28	156.5	142.4	89.8	76.9	78.3	85.1	67.6
U29	105.3	142.8	93.5	74.9	74.8	83.6	67.9
G30	na	137.5	92.7				
G31	na	136.1	92.5				
A32	153.6	138.8	92.4	75.9	72.5	81.9	64.8
G33	na	135.0	92.3				65.8
C34	97.9	139.8	94.1			82.0	65.1
A35	154.2	139.3	92.7	76.2	73.0	82.4	65.2
C36	97.6	141.8	92.4			84.1	65.7

[a]Not applicable

***Table A.VI**. Structural statistics of the final ensemble of 15 structures.*

Distance restraints	
Intranucleotide	294
Internucleotide	171
Hydrogen bonds	34
Non-NOE	48
Rms deviations (Å, °)	
Distance restraints (547)	0.075±0.002
Dihedral restraints (272)	1.203±0.050
Rms deviation from idealized geometry	
Bonds (Å)	0.0125±0.0004
Angles (°)	1.998±0.092
Impropers (°)	0.471±0.023
Restraint violations	
Number of distance violations[†] > 0.4Å	1±1
Number of dihedral violations[‡] > 4°	8±2
Atomic rms deviations (Å)[¶]	1.72±0.41

[†]None larger than 0.5Å. [‡]None larger than 6°. [¶]With respect to the average structure.

Summary

&

Samenvatting

Summary

Recent developments in NMR spectroscopy and X-ray crystallography have resulted in a large dataset of remarkable RNA structures. Notwithstanding that RNA has only four different building blocks, a surprisingly large variety of different structural motifs are found in RNA. The structure elucidation yielded a better understanding of the folding principles and functional mechanisms of these molecules. For instance, the analysis of tRNA has led to a better insight in the codon/anti-codon interaction with the mRNA. In this thesis, the results of structural NMR studies on two functionally different RNA fragments, the substrate hairpin of the *Neurospora* VS ribozyme and the SRV-1 pseudoknot involved in ribosomal frameshifting, are described.

An overview of the large variety of structural interactions and motifs that have been found in RNA structures is presented in Chapter I. In general, RNA consists of helical regions intersected by single stranded regions, loops and bulges that interact to a large extent with other domains in the molecule, thereby determining its overall fold. Similarly to α-helices and β-sheets in proteins the complex folds of RNA molecules contain reoccurring motifs that stabilize tertiary structure elements, e.g. base zippers, A-minor motifs, adenosine platforms, U-, reversed U-, S- and C-turn motifs. The first structure resolved at high-resolution, the structure of tRNA, already revealed a large number of these recurring themes. However, being the only structure solved at that time these motifs could not be recognized as such.

In the *Neurospora* VS ribozyme, a long-range interaction is important for folding this RNA molecule into the active conformation. Thus, in the presence of magnesium a pseudoknot-like interaction is formed between the hairpin loop of the substrate part and a hairpin loop of the remaining part. Chapter II describes the NMR structure determination of the internal loop of the substrate hairpin of the *Neurospora* VS ribozyme. This internal loop comprises the actual cleavage site. It turned out that the guanosine and the adenosine on either side of the scissile phosphate are both involved in a tandem sheared GA base pair. The internal loop additionally holds an $A^{+}C$ base pair, that is isosteric to the GU base pair, with the adenosine base protonated at the N1 position. The structure showed that for the cleavage site the orientation of the scissile phosphate group with respect to the oxygen of the hydroxyl group was not compatible with the proposed cleavage intermediate for an in-line attack. Also the pH effects on the chemical shift of the C2 nucleus of A7 that we

observed could not be directly correlated with the biochemical analysis of the pH-dependency on the cleavage rate. Thus, we concluded that the NMR derived structure represents some kind of ground state and that prior to cleavage a significant conformational change will occur.

Chapter III presents an overview of the different pseudoknot structures that have been solved by X-ray crystallography and NMR spectroscopy. The pseudoknot structure was first postulated for and found in the 3'-terminus of the TYMV RNA genome. It is a specific folding motif in which a single stranded tail of a hairpin structure forms Watson-Crick base pairs with bases of the loop. Pseudoknots are present in practically all classes of RNA and are involved in many biological processes, such as replication and translation. First the specific fold of H-type pseudoknots is described using the classic structure of the 44n fragment of TYMV, which shows several interesting structural features. Two stems stack in a coaxial manner accompanied by an increased helical twist between the base pairs at the junction. Furthermore, the overall structure is stabilized by stem-loop interactions. Similar structural features are observed in other RNA pseudoknots as well. The pseudoknots in BWYV, MMTV and SRV-1 promote ribosomal frameshifting, a regulatory mechanism used for the tuning of the expression of the appropriate amount of proteins that are translated from the overlapping region of the polycistronic viral mRNA. The bent conformation of the two stems and the stem-loop interactions present in the BWYV and MMTV pseudoknots, have been postulated to be important for the frameshift mechanism. However, despite the structural insight obtained for these pseudoknots it is still not clear how they function in the ribosomal frameshifting. The HDV ribozyme, a double-nested pseudoknot, differs from the above mentioned classical H-type pseudoknots. The different loop regions of the first pseudoknot are relatively large and have secondary structure elements that form a second pseudoknot interaction. Therefore it was called a double-nested pseudoknot. These pseudoknot interactions are structurally relevant since they determine the overall fold of the ribozyme by bringing different domains together. Since they involve long-range interactions we prefer to call them tertiary rather than pseudoknot interactions.

One of the new developments in NMR spectroscopy is the usage of dipolar couplings obtained by alignment of nucleic acids and proteins in anisotropic solutions containing bicells or filamentous phages. Dipolar couplings between covalently bound nuclei can be translated into the relative orientations of for instance ^{1}H-^{13}C and ^{13}C-^{13}C

vectors and therefore hold long-range angular information. In particular structure determination of RNA molecules can benefit from this approach, since structure calculations of RNA molecules mainly depend on local sequential restraints obtained by the traditional NOE and J-coupling data. Chapter IV describes two experiments used for the assignment of all carbon and proton resonances within the adenosine base ring system, which is the first step in collecting one-bonded dipolar coupling constants. In addition to a full assignment of the carbon and proton resonances, one of the two experiments unambiguously correlates the adenine H8 and H2 resonances, using an alternative magnetization transfer pathway compared to previously published experiments. The experiments were successfully applied to the SRV-1 frameshifting pseudoknot. With the assignment of the H2 resonances a large number of distance restraints to other protons could be obtained from the NOESY spectra, which turned out to be of great value in the structure determination.

The solution structure of the SRV-1 pseudoknot, described in Chapter V, turned out the to be highly ordered. It comprises two stem regions of six base pairs each that coaxially stack, similar to what has been observed for the TYMV and T2 pseudoknots. Loop 1 contains a single nucleotide that crosses the major groove of stem 2, while the minor groove of stem 2 is spanned by a total of nine residues of loop 2. Six out of these nine residues are stacked, stabilized by three base triple interactions, and the cytosine in the middle of the loop forms a bulge. A single ribose zipper motif was found between A26 in loop 2 and C16 of stem 1. Different from the MMTV and TYMV pseudoknot, here the last residue at the 3'-end of loop 2 (A28) is involved in a hydrogen bond interaction with the hydroxyl group of G14 in stem 1. The sugar of A28 is in the 2'-endo conformation and at this point the backbone changes in orientation by a so-called S-turn enabling the formation of the U-A base pair at the junction. Several results of the biochemical data previously obtained can be explained by the structure and it appears that stem-loop interactions are of great importance in the actual frameshift mechanism.

The studies described in this thesis provide a basis for a further understanding of RNA folding. In addition, the structure of the substrate hairpin of *Neurospora* VS ribozyme extends our present knowledge on internal loops containing G-A base pairs. The structure of the SRV-1 RNA pseudoknot enlarged greatly our insight of H-type pseudoknot

architecture and provided important leads for unraveling the ribosomal frameshift mechanism.

Samenvatting

In de afgelopen jaren hebben de ontwikkelingen op het gebied van de NMR-spectroscopie en Röngten-kristallografie geleid tot een groot aantal nieuwe opgeloste RNA structuren. Ondanks het feit dat RNA is opgebouwd uit slechts vier verschillende bouwstenen, is er een grote diversiteit aan structurele motieven gevonden. Deze hebben geleid tot een beter begrip van de vouwingsprincipes en werkingsmechanismen van deze nucleïnezuren. De studies aan tRNA hebben bijvoorbeeld geleid tot een beter inzicht in hoe de codon/anti-codon-interactie met het mRNA plaatsvinden. In dit proefschrift zijn de resultaten beschreven van het structuur onderzoek met behulp van NMR-spectroscopie aan twee verschillende RNA fragmenten, de substraat-haarspeld uit het *Neurospora* VS ribozym en de SRV-1 pseudoknoop betrokken bij ribosomale leesraamverschuiving.

De inleiding van dit proefschrift geeft een overzicht van de structurele interacties en verschillende motieven die men tot nu toe in RNA-molekulen heeft gevonden. In het algemeen is RNA opgebouwd uit helices met daartussen enkelstrengs-gebieden, lussen en "bulges" die een interactie kunnen aangaan met andere delen van het molekuul en daarmee de algehele vouwing bepalen. Net als de α-helix en β-sheet in eiwitten, treffen we in de complexe RNA-molekulen eveneens verschillende terugkerende motieven aan die de drie-dimensionale struktuur stabilizeren. Voorbeelden hiervan zijn de zogenaamde 'base zippers', het 'A-minor' motief, het adenosine-plateau en de zogenaamde 'U-, reverse U-, C- en S-turn' motieven. De eerste hoge-resolutie RNA struktuur, de struktuur van tRNA, bevatte al een groot aantal van deze terugkerende motieven, maar deze konden uiteraard niet als zodanig worden herkend aangezien tot op dat moment de tRNA-struktuur de enige opgeloste RNA struktuur was.

In het *Neurospora* VS RNA ribozym is er een lange-afstand-interactie belangrijk bij de algehele vouwing van de actieve conformatie. In aanwezigheid van magnesium wordt een pseudoknoop-achtige interactie gevormd tussen de haarspeldlus van het substraat-gedeelte en een haarspeldlus van het overige deel van het ribozym. Hoofdstuk II beschrijft de NMR-struktuurbepaling van de interne lus van de substraat-haarspeld van het *Neurospora* VS ribozym. Deze interne lus van het substraat bevat de knipplaats. Uit deze studie bleek dat de guanosine en de adenosine aan weerszijde van de knipplaats ieder betrokken zijn bij een z.g. tandem G.A base-paar. De interne lus bevat tevens een $A^{+}C$ base paar, waardoor het isosteer is aan het G.U base-paar. In dit base-paar is de adenine-base

geprotoneerd op de N1 positie. Uit de struktuur blijkt dat de oriëntatie van de fosfaat groep ten opzichte van de zuurstof van de hydroxylgroep niet direct kon leiden tot het intermediair dat voor de klievingsreaktie is voorgesteld. Ook de door ons waargenomen pH-effecten op de chemische verschuiving van de C2 kern van A7, zijn niet in overeenstemming met de biochemische analyses naar de pH-afhankelijkheid op de snelheid van klieving. Hieruit werd geconcludeerd dat de uit NMR-gegevens bepaalde struktuur een soort van grond-toestand is en dat kort voor de klievingsreaktie van het RNA een significante strukturele verandering moet plaatsvinden.

Hoofdstuk III geeft een overzicht van de RNA-pseudoknoop-strukturen die zijn opgelost met NMR-spectroscopie en Röntgen-kristallografie. De pseudoknoop struktuur was voor het eerst gepostuleerd voor en herkend in het 3'-uiteinde van het TYMV RNA genoom. Het is een specifiek vouwingsmotief waarbij een enkelstrengs uiteinde van een haarspeldstruktuur Watson-Crick base-paren vormt met de basen uit de lus. Pseudoknopen komen in bijna iedere klasse van RNA-molekulen voor en zijn betrokken in vele biologische processen, zoals replicatie en translatie. Allereerst wordt in dit hoofdstuk de specifieke vouwing van H-type pseudoknopen beschreven aan de hand van de klassieke struktuur van het 44n-fragment uit TYMV RNA, dat verschillende interessante strukturele elementen bevat. De twee stam-gedeelten zijn coaxiaalsgewijs gestapeld. Dit gaat gepaard met een ontwinding van de totale helix op het knooppunt waar alle strengen bij elkaar komen. Daarnaast wordt de algehele struktuur ook nog gestabiliseerd door stam-lus interacties. Vergelijkbare structurele elementen zijn eveneens waargenomen in andere RNA pseudoknopen. De pseudoknopen in BWYV, MMTV en SRV-1 veroorzaken ribosomale leesraamverschuiving, een regulatie-mechanisme voor de synthese van de juiste verhouding van eiwitten die getransleerd worden van overlappende genen in het virale mRNA. De hoek tussen de twee stam-gedeeltes en de loop-stam interacties, aanwezig in de strukturen van de pseudoknopen van BWYV en MMTV, zouden belangrijk kunnen zijn in het mechanisme voor ribosomale leesraamverschuiving. Ondanks deze verworven strukturele inzichten, kan men voor de MMTV en de BWYV pseudoknoop toch geen duidelijk antwoord geven op de vraag hoe zij functioneren als regulerend element in ribosomale leesraamverschuiving. Het HDV ribozym, een "double-nested" pseudoknoop, verschilt van de voorgaande H-type pseudoknopen. De lus-gedeelten van de eerste pseudoknoop zijn relatief groot en bevatten op zichzelf weer secundaire struktuurelementen die interacties aangaan waardoor een

tweede pseudoknoop wordt gevormd. Daarom wordt het een dubbel-genestelde pseudoknoop genoemd. Deze pseudoknoop-interacties bepalen in het geval van HDV zeer duidelijk de algehele vouwing van het ribozym. Aangezien we hier spreken over lange-afstand-interacties, zouden deze beter als tertiaire interacties aangemerkt kunnen worden dan als pseudoknoop-interacties.

Een van de nieuwe ontwikkelingen in de NMR-spectroscopie is het gebruik van dipolaire koppelingen door middel van de uitrichting van nucleïnezuren en eiwitten in anisotropische oplossing die bijvoorbeeld bicellen of filamenteuze fagen bevatten. Dipolaire koppelingen tussen covalent gebonden kernen kunnen worden vertaald in relatieve orientaties van bijvoorbeeld ^{1}H-^{13}C en ^{13}C-^{13}C vectoren en bevatten dus lange-afstand-informatie. Juist de struktuurbepaling van RNA-molekulen heeft profijt van deze informatie, omdat de struktuurberekening van RNA-molekulen voor het grootste gedeelte afhankelijk is van sequentiele afstand- en torsiehoek-informatie verkregen uit NOE en J-koppeling data. Hoofdstuk 4 beschrijft twee experimenten voor de toekenning van alle proton- en koolstofresonanties uit het adenosine-base-ring-systeem. Dit is de eerste stap in het verkrijgen van vector-informatie uit één-bands dipolaire koppelingen. Naast de volledige toekenning van koolstof- en protonresonanties, kan met behulp van een van de experimenten ook de H2 met de H8 ondubbelzinnig gecorreleerd worden. De overdracht van magnetisatie in dit experiment is verschillend van eerder gepubliceerde H2-H8 correlatie-experimenten. De experimenten zijn met succes toegepast op de pseudoknoop, die leesraamverschuiving veroorzaakt in SRV-1 RNA. Met de toekenning van de H2's werd uit het NOESY-spectrum een grote hoeveelheid afstand-informatie naar andere protonen verkregen, die belangrijk bleek te zijn bij de struktuurbepaling van de pseudoknoop.

De struktuur van de SRV-1 pseudoknoop in oplossing wordt beschreven in Hoofdstuk 5. Hij blijkt bijzonder goed gedefinieerd te zijn. De struktuur bestaat uit twee (elk uit zes base-paren bestaande) stam-gedeeltes, die vergelijkbaar aan de TYMV- en T2-pseudoknoop coaxiaal doorstapelen. Lus 1, bestaande uit één residu, overspant de diepe grote groeve van stam 2, terwijl de ondiepe kleine groeve van stam 1 wordt overbrugd door de negen residuen van lus 2. Zes van deze laatstgenoemde negen residuen stapelen opeenvolgend door, gestabiliseerd door drie "base triple"-interacties, waarbij de cytosine in het midden van de lus niet betrokken is in de doorstapeling. Daarnaast werd ook een

'ribose-zipper' motief aangetroffen tussen A26 uit lus 2 en C16 uit stam 1. Verschillend van de MMTV en TYMV pseudoknoop vormt het laatste adenine residu (A28) aan het 3'-uiteinde van lus 2 een waterstofbrug met een hydroxylgroep van G14 uit stam 1. De suiker-conformatie van deze adenine (A28) is 2'-endo en verandert de richting van de "backbone" zodanig dat middels een S-bocht de formatie van het U-A base-paar op de kruising van beide stam-gedeeltes mogelijk is. Eerder uit biochemische experimenten verzamelde resultaten kunnen worden verklaard aan de hand van de nieuwe struktuur waaruit blijkt dat de stam-lus interacties een niet onbelangrijke rol spelen in het mechanisme van leesraamverschuiving.

De studies zoals beschreven in dit proefschrift hebben bijgedragen aan een beter begrip wat betreft RNA-vouwing. Daarnaast heeft de struktuur van de substraat-haarspeld uit *Neurospora* VS onze kennis over interne lussen in RNA, die G-A base-paren bevatten, verdiept. De struktuur van de SRV-1 RNA pseudoknoop heeft de inzichten vergroot in de architectuur van H-type pseudoknopen en heeft aanwijzingen opgeleverd die zullen leiden tot een beter begrip van het mechanisme van leesraamverschuiving tijdens translatie.

Curriculum Vitae

De auteur van dit proefschrift werd geboren in Boxmeer op 8 juni 1969. Hij behaalde zijn VWO diploma aan het Titus Brandsma Lyceum te Oss in 1987. Hierna studeerde hij Scheikunde aan de Katholieke Universiteit Nijmegen. Hierbij werd in het kader van een bijvak een onderzoeksstage verricht bij de vakgroep Biochemie van de medische faculteit onder leiding van Dr. G.J. Bosman en Prof. dr. W. de Grip, een extra bijvakstage Biochemie aan de Universiteit van Colorado in de Verenigde Staten onder leiding van Prof. dr. A. Pardi, en een hoofdvakstage in de vakgroep Biofysische Chemie onder leiding van Prof. dr. C.W. Hilbers en Dr. H.A. Heus. Het doctoraal examen werd in 1993 afgelegd.

In januari 1994 trad de auteur als onderzoeker in opleiding in dienst van Nederlandse Organisatie voor Wetenschappelijk Onderzoek. Het onderzoek dat is beschreven in dit proefschrift werd verricht in het kader van een samenwerkingsverband tussen de groep Genexpress van het Leids Instituut voor Chemisch onderzoek van de Universiteit van Leiden onder leiding van Prof. dr. C.W.A. Pleij en de vakgroep Biofysische Chemie, een onderdeel van de onderzoeksschool het NSR center, aan de Katholieke Universiteit Nijmegen onder leiding van Prof. dr. C.W. Hilbers. Per 1 november 1998 kwam de auteur in dienst van de Katholieke Universiteit Nijmegen als junior researcher.

Nawoord

Zelden gaan de dingen vanzelf. Soms moet je beseffen dat je niet alles alleen kunt of hoeft te doen, dat is het mooie aan het werken in teamverband. Tijdens de totstandkoming van dit proefschrift heb ik dan ook met een groot aantal mensen heel prettig (, direkt of indirekt,) samengewerkt en daarvoor ben ik ze allen zeer dankbaar. Allereerst is dat professor Hilbers, die mij de mogelijkheid heeft gegeven dit onderzoek te kunnen doen en wiens discussies meestal hele nieuwe inzichten op zaken konden geven. Er zijn zogezegd, flink wat beren van de weg gehaald. Ook wil ik graag professor Pleij van de Universiteit van Leiden, in wiens groep ik een deel van mijn onderzoek heb opgestart, bedanken. Ik ben daar als zuiderling bijzonder goed terecht gekomen, mede door de gemoedelijke (zuidelijke) sfeer die hij daar weet te scheppen. Zeker niet op de allerlaatste plaats wil ik Hans Heus bedanken voor zijn hulp en adviezen. En niet te vergeten, mooi schrijven is een kunst. Dankzij Hans heeft al het goede werk een extra mooi jasje gekregen.

Zoals al gezegd heb ik mijn promotie-tijd zowel in Leiden als in Nijmegen doorgebracht en op beide plaatsen was er naast het werk ook tijd voor een rustig moment, voor een 'bakske' koffie. Alle mensen waarmee ik in Leiden *"een bakkie pleur deed"*; Paul II, Anne-Marie, Maarten, René, Lian, Barend, Jan, Koen, Erik, Anneke, Tineke, Gerrit, Jules, Volkert, Raymon, Sjoerd, Hugo en in het bijzonder Birgit en Dico, mijn beste maatjes, wil ik bedanken voor de fijne twee jaar daar. Tuurlijk zijn er grapjes gemaakt over mijn zachte G, maar dat vergeef ik ze!

In Nijmegen was ik natuurlijk al thuis. De koffietafel daar wil ik ook graag bedanken voor alles, beginnende vanaf het hoekje onder de bananenboom (inmiddels palm-boom); Nelly, Jan, Aafke, Ben, Ine, Arno, van Os, (Nachtegaal oh nee,) Gerda, Hans, Gerrit, Harry, Carla, Hanny, Maaike, Rieko, Hartmut, Adrie, Dinu, Adriana, Paul. Verder ook alle promovendi, studenten en postdocs die ik heb meegemaakt, voor onder andere de gezellige etentjes; Michaël, Rutger, Tineke, Lennard, Janine, Chris, Henno, Maria, Ippel, Jos, Cathelijne, Antonio, Alfonso, Hugo, Sander, Jaap, Tine, Christophe, Chiara and Tim.

Een paar mensen verdienen nog extra aandacht omdat ze mij meer dan gemiddeld hebben geholpen met ondezoek en aanverwante dingen. Met Rob en Caroline heb ik heel erg leuk samen kunnen werken aan het *Neurospora* project. Paul en Alexandra ben ik heel dankbaar

voor het natte werk dat ze me uit handen hebben genomen toen ik daar weinig tijd voor had. Jan bedankt voor alle T7 RNA polymerase zuiveringen, Babs en John voor al het papierwerk van declaraties en bestellingen, tot brieven voor beurzen, en Jos bedankt dat je altijd klaar stond als er weer iets was met een computer of printer. Marco, Geerten en Michaël wil ik ook nog extra bedanken voor al hun hulp achter de spectrometer.

Naast alle dankwoorden aan mijn collega's zijn er natuurlijk ook mijn vrienden en familie die mij altijd enorm gesteund hebben. Om ze allemaal te noemen zou te veel zijn, maar ik kan niet om de heren van NVAA (Heimen, Jos, Paul, Bert en Peter) heen. Onze lijfspreuk “niets is moelijk voor hen die werkelijk willen” is ook zeer toepasselijk geweest voor mijn onderzoek. In de tijden dat ik niet verwachtte dat het zou lukken om de struktuur van die verflickste pseudoknoop op te lossen heb ik aan deze spreuk gedacht. En daarmee is het me dan toch maar mooi gelukt en ben ik ook het meest trots op Hoofdstuk 5, temeer omdat de ‘concurenten’ in Amerika het opgegeven hebben om eraan te werken. Tenslotte mijn ouders en mijn broer, dat weet ik, zijn heel erg trots op me en dat heeft ook moed en inspiratie gegeven de afgelopen jaren om het tot een mooi resultaat te brengen.